(Mis)perceiving Infection

by

Nicholas M. Michalak

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2020

Doctoral Committee:

Associate Professor Joshua M. Ackerman, Chair
Emeritus Professor Phoebe Ellsworth
Professor Richard Gonzalez
Assistant Professor Arnold Ho
Assistant Professor Neil Lewis, Jr., Cornell University

Nicholas M. Michalak

nickmm@umich.edu

ORCID ID: 0000-0001-9122-7291

Dedication

To Mom and Dad, who have sacrificed and loved and supported me throughout a gajillion years of schooling. To my sister Alex, who has a scientific understanding of how full of shit I am and yet loves and supports me anyway. To Mark, my brother in all the ways that matter, who helped me find direction and purpose after college and who will welcome me into his research institute and commune when I have nowhere else to go. And to Kelly, the love of my life who inspires me every day to empathize with people, ignore charlatans, study numbers, and appreciate nature.

Acknowledgements

I'm grateful for a variety of people who made it possible for me to start my PhD program at the University of Michigan and complete it successfully and happily. First, I want to thank the group of people who supported me while I applied to graduate school one, two, three times. This group includes Drs. Britain Scott, Elise Amel, Dalma Martinovic, Ethan Young, Chloe Huelsnitz, Stephanie Cantu, Allie Farrell, Jeff Simpson, Vladas Griskevicius, and Kathleen Vohs as well as Payman Saghafi and Bob Ottmon. Drs. Britain Scott, Elise Amel, Dalma and Martinovic inspired my interests in science, biology, psychology, and research, and encouraged me to apply to graduate school. Drs. Stephanie Cantu and Allie Farrell took a chance on me as a college graduate looking for research experience, and they mentored me as a budding evolutionary social psychologist and supported me while I applied to graduate school. I joined an undergraduate lab research team with Drs. Ethan Young and Chloe Huelsnitz who supported me while they were also applying to graduate school at the University of Minnesota. Drs. Jeff Simpson and Vladas Griskevicius welcomed me not only into their research labs but as an auditor into their graduate seminar on evolutionary social psychology. If Vlad had not told me two weeks before the deadline that Dr. Ackerman was looking for students, I would never have applied to the University of Michigan. Dr. Kathleen Vohs gave me the keys to her lab for two years as her lab manager and afforded me the opportunity to develop the final skills and accumulate the experience necessary to compete as a graduate school applicant. Payman Saghafi helped me study for the GRE and mentored and paid me as a tutor while I was a barista, research assistant, and lab manager. Bob Ottmon lent me a valuable set of eyes and perspective for my

mountains of personal statement drafts. Tom Durfee, Natalie Loots, Lauren Graff, Eric Swenson, Katy Cassingham, Charlie Henke, Saiid Lewis, and Adam Wozniak kept my life silly, friend-full, and adventurous while I juggled random jobs as a barista and a tutor in St. Paul/Minneapolis. Speaking of my barista job: My St. Paul Espresso Royale crew members were basically family for me after college while I tried to get into graduate school. I've listed these wonderful people as persons who helped me out in one way or another, but they are all simply friendly and giving people, many of whom I've shared one or several drinks or adventures with.

During graduate school, I had the good fortune to befriend a group of brilliant and down to earth graduate students, some at Michigan but others at universities like Pennsylvania that are almost as nice as Michigan. The first of many of these fortunes was starting the social psychology program at Michigan with the social-est of cohorts: Koji Takahashi, Iris Wang, Dr. Todd Chan, and Dr. Kaidi Wu. I spent many late nights talking about statistics, politics, and life with my Bestminster roommate Koji. I dug deep into messy data, picked berries, water-colored, and data scienced with Iris. I burned hard-earned graduate stipends and brain cells drinking expensive whiskey and cheap French fries with Todd. And I spent many puzzled moments watching Kaidi give research presentations but also many amazed moments watching her speak at length about a variety of socio-cultural phenomena. I shared an office as well as seemingly endless (at times unsavory) banter and research discussions with Dr. Darwin Guevarra, Dr. Steve Tompson, Nadia Vossoughi, and Zachary Reese. I ran easily hundreds of miles and ate dozens of post-run brunches on the weekends with Dr. Mike Hall. I had even more deep stats, coding, and open science discussions (sometimes in the Netherlands) with Aki Gormezano. Kris Smith, who unfortunately, had to study psychology in Philadelphia, kept me on my toes during discussions about evolution, psychology, statistics, and twitter trolls at SPSP and HBES. And I had so many

more insightful discussions about too many topics and with too many people to name (but I'll try): Dr. Izzy Gainsberg, Rachel Fine, Dr. Sara Chadwick, Dr. Ariana Orvell, Dr. Veronica Derricks, Rebecca Marks, Dr. Sarah Westrick, Cristina Salvador, Wilson Merrell, and Soyeon Choi.

I was extremely fortunate to find wise, kind, and generous mentors while at Michigan. Dr. Neil Lewis, Jr. green-lite my application materials for Michigan; recommended me for my departmental statistics consultant role; teamed-up with me on our now registered report of a cross-temporal meta-analysis on stereotype threat; offered me a postdoc and served on my committee. Not only was Neil a mentor, he inspired and encouraged me to care about open science, real-world applications for psychology, and the importance of not being a dick. Dr. Aaron Weidman encouraged me to focus on the trade-offs between the nuances of statistical decisions and the robust, easy-to-explain, qualitative takeaways. Of course, we also had rich discussions about methods, statistics, and the Academy. Dr. Oliver Sng somehow managed to impart on me the wisdom of an Emeritus professor through the eyes of an up-and-coming researcher. Oliver inspired me through his scientific curiosity and careful words. Dr. Fred Conrad mentored me in survey methodology, and he inspired me not only to think deeply about how we ask questions in surveys but also how participants experience those surveys. Dr. Arnold Ho introduced me to the science of social inequality and prejudice, topics I had too little exposure to coming into a graduate program in social psychology! Dr. Adrienne Beltz and Dr. Rich Gonzalez had complimentary influences on how I approach data and statistics. Among too many topics to list here, Rich passed on a zeal for contrasts and a distaste for omnibus tests; Adrienne imparted a love for maximum likelihood and models of individual differences. Dr. Phoebe Ellsworth motivated me to ask questions people care about, design better studies, and

notice snake-oil for what it is. Over the course of this incredible program, I've been blessed with the opportunity to discuss wide-range social science topics with some of the most friendly and brilliant scholars on the planet: Drs. Robin Edelsetin, Allison Earl, Denise Sekaquaptewa, Bill Chopik, David Dunning, Oscar Ybarra, Matthew Diemer, Ethan Kross, Shinobu Kityama, and Sari van Anders.

I'm giving Dr. Josh Ackerman his own paragraph. Josh has influenced nearly every part of my progress as a doctoral student and my growth as a social scientist and a professional. We planned studies, built presentations, and wrote papers together. We discussed theory, replication, and generalizability, and we discussed career strategies and productivity hacks. Josh was easy to access, generous with his time, clear about goals, cool under pressure, understanding and forward looking when things didn't work out. Josh was more than someone to talk with about science and publications. Josh took an active role in my growth and success as a professional. One day I hope to be half as scientifically balanced, reliable, organized, kind, and generous as Josh Ackerman. My future is bright due in no small part to mentorship.

To sum it all up, I'm filled with gratitude. I'm fortunate to have such a long list of incredible people to thank. I'm also fortunate to have studied in the Michigan Psychology Department, held together largely by the tireless staff at Student Academic Affairs. I'm also fortunate to have enjoyed generous stipends, fellowships, health care plans, selfless graduate student union (GEO) leaders, and all the other affordances made possible by the Rackham Graduate School. Not everyone is lucky enough to enjoy the freedom, resources, good company, and supportive mentors that I did during my time at Michigan. I hope I can pass it forward.

Table of Contents

List of Figures

List of Tables

Abstract

How do people detect whether someone else poses an infection risk? Over the course of evolutionary time, humans evolved sophisticated physical and psychological machinery for detecting and reducing fitness costs associated with infectious pathogens. Whereas the physical immune system evolved to detect invasive pathogens within the body and eliminate them, the "behavioral immune system" evolved to detect infection risks outside the body and engage cognitive, emotional, and behavioral mechanisms that serve to prevent infectious pathogens from entering the body in the first place. My dissertation research focuses on the psychological mechanisms that enable people to detect infection risks. Importantly, such mechanisms entail a sort of functional misperception of infection risks: given uncertainty in detecting infection and asymmetrical costs associated with false positive and false negative errors, the behavioral immune system perceives both objectively diagnostic infection cues as well as physically anomalous yet benign cues as indicators of infection risk. In other words, in order to avoid the high cost of ignoring a true infection risk, the behavioral immune system is biased to perceive infection risk in anomalous cues, even ones that are objectively benign. After I review this hypothesis and its empirical tests in Chapter 1, in Chapters 2 and 3 I report two series of studies—one series per chapter— designed to discover novel psychological mechanisms of the behavioral immune system. In Chapter 2, I employ the Implicit Association Test to explore whether people associate infection more strongly with one benign, anomalous cue (facial disfigurement) than with another (obesity). Across four studies, I find evidence that it does. Then, in Chapter 3, I employ trait-listing, drawing, and reverse correlation methods to estimate

how people mentally represent infected others. I find that people list traits and draw pictures of infected people with mostly infection cues. But when people complete a reverse correlation task, they generate infected people with a mixture of threat cues. I conclude my dissertation with future directions for behavioral immune system research.
Keywords: behavioral immune system, threat management, evolutionary psychology, social perception, mental representations

Introduction

Infectious disease has been a recurring source of mortality throughout our evolutionary history, and even to this day, it continues to threaten lives across the world, especially the lives of young children and older adults in poor countries. So, it is not surprising that humans have, over evolutionary time, developed sophisticated physical machinery—an immune system—for identifying invasive pathogens and eliminating them. Of course, the immune system costs valuable energy, so completely ignoring harmful pathogens until they enter the body wastes energy that could be used for other fitness-enhancing activities. To reduce such costs, humans have likely developed psychological and behavioral mechanisms—a behavioral immune system—for identifying and mitigating pathogen infection risks.

In my dissertation, I focus on how such a behavioral immune system identifies infection risks posed by other people. In chapter one, I start by reviewing key assumptions of the behavioral immune system as well as empirical research on how it infers whether others pose an infection risk. In the following chapters, I elaborate on the social-cognitive mechanisms that aid in detecting infection risks posed by others, and I report studies designed to test some of those mechanisms. Specifically, in chapter two, I elaborate on the kinds of cues the behavioral immune system evolved to process—its proper domain of cues—and the kinds of cues it is able to process, even if it did not evolve to do so—its actual domain of cues. Then I report a series of studies designed to test whether people more strongly associate infection concepts with cues that resemble theoretically proper domain cues (e.g., rashes) than with cues that less closely resemble such proper domain cues (e.g., obesity). Finally, in chapter three, I elaborate on how people may

generate mental images of infected others that appear differently from a simple sum of theoretical infection cues (i.e., rashes + pustules + asymmetry + swelling + runny nose) or from an indiscriminately negative person. Then I report a series of studies designed to estimate how people mentally represent what an infectious person looks like. Finally, I conclude my dissertation with recommendations for future research that could address open questions in the pathogen avoidance psychology literature as well as the threat management literature more broadly.

Chapter I. The Behavioral Immune System

The specter of disease caused by infectious pathogens has imposed an immense selection pressure on the evolution of life for billions of years. Human life is no exception. In fact, most human lives have ended at the hands of pathogenic infections, and new infections continue to cause many millions of deaths worldwide (Anderson & May, 1992; Dobson & Carper, 1996; Wolfe et al., 2007; World Health Organization, 2018). To combat this ominous threat posed by infectious pathogens, humans have evolved a physiological immune system designed to distinguish invading pathogens from the body's own cells and mobilize physiological defenses to reduce infection. They also appear to have evolved a "behavioral immune system," a suite of psychological and behavioral mechanisms designed to detect the presence of pathogens and engage appropriate cognitive, emotional, and behavioral responses to avoid them (Ackerman et al., 2018; Curtis et al., 2004; Oaten et al., 2009; Schaller, 2016; Tybur et al., 2013; Tybur & Lieberman, 2016).

The behavioral immune system cannot detect infectious pathogens directly. Instead, it infers their presence using cues that historically have correlated with pathogenic infection. These cues are often noticeable in appearance, as infection can cause rashes, lesions, discoloration, swelling, bleeding, and/or nasal discharge, among other physical changes. Of course, physical features also may resemble these cues without connoting the presence of infection. For example, rashes may occur due to skin irritation, or noses may run because of cold ambient temperatures. The similarity between true and false infection cues, in combination with the imperceptibility of pathogenic agents, produces a signal detection problem in which identification errors are

virtually certain. Further, the costs of different forms of error are unequal. Failure to identify a dangerous pathogen's presence (e.g., mistaking swelling from deadly Diphtheria as benign) is far costlier to fitness than mistakenly inferring a parasitic threat that is not there (e.g., mistaking a birthmark as an infectious rash). To cut through the noise, selection appears to have favored a cognitive bias that reduces fitness costs by over-perceiving pathogen threat from imperfect cues (Haselton et al., 2015; Haselton & Buss, 2000; Haselton & Nettle, 2006; McKay & Efferson, 2010; Nesse, 2005).

Researchers who study pathogen avoidance psychology have used this adaptive over-perception to partly explain stigma directed toward people with anomalous yet benign facial features such as obesity, port-wine stain birthmarks, strabismus (i.e., crossed-eyes), old age, and physical disabilities (Ackerman et al., 2009; Kurzban & Leary, 2001; Lieberman et al., 2012; Miller & Maner, 2012; Park et al., 2003; Ryan et al., 2012; Schaller & Neuberg, 2012; van Leeuwen et al., 2015). To understand the behavioral immune system's tendency to perceive infection in non-infectious features, researchers in this area assume that the behavioral immune system is a functional module of the mind—it comprises cognitive mechanisms with specialized functions with restricted inputs (Barrett, 2012; Barrett & Kurzban, 2006; Schaller, 2016; Sperber, 1994; Sperber & Hirschfeld, 2004). In the context of pathogen threat, the function of the behavioral immune system is to process inputs reflecting true signs of infection (constituting the "proper" domain for this system). Yet, because of the signal detection problem described above, the design elements of this system also process cues that share perceptual properties with cues from the proper domain. The combined set of these cues constitute the actual domain of the behavioral immune system. So, theoretically, the behavioral immune system should react to true

infection cues as well as false infection cues if such false infection cues share perceptual properties of true infection cues.

Researchers have tested this over-perception hypothesis using a variety of methods. For example, in one analysis, Ackerman et al. (2009) found that people participating in a dot-probe task who were psychologically primed with pathogen threat looked longer at faces with benign birthmarks or crossed-eyes than at faces without those features. This suggests people were attending more to anomalous yet benign facial features when pathogen infection was salient as if those features indicated infection risk. In another study, people expressed more disgust in response to, and were less willing to touch, objects that had been touched by people with visible cues of influenza compared to the same objects that had been touched by visibly healthy targets (Ryan et al., 2012). People also expressed more disgust by and less willingness to touch those objects when the target people bore non-infectious facial blotches, suggesting participants perceived features that merely resembled infection cues as if they were true indicators; that perception made them avoid touching objects that had been "contaminated" through physical contact. As a third and final example, people for whom pathogen threat was salient recorded stronger automatic negative associations with obese targets than people in a control condition (Park et al., 2007) and people who reported higher concerns about or disgust from pathogenic objects (e.g., feces, bloody cuts, chewed pencils) tended to report stronger negative attitudes toward obese people (Lieberman et al., 2012; Park et al., 2007). All together, these results support the hypothesis that people misperceive infection risk in benign features that share perceptual properties with true infection indicators (e.g., rashes, swelling).

Researchers have extended this over-perception hypothesis to include cues that less plausibly resemble true infection indicators but nonetheless could be associated with harmful

pathogens. The most prominent example is foreign appearance perceived in ethnic outgroups. A sizeable body of evidence suggests people under pathogen threat—either people high in pathogen threat concern or people exposed to pathogenic materials in a laboratory—report stronger negative attitudes toward ethnic outgroups (i.e., anti-immigrant attitudes) (Aarøe et al., 2017; Faulkner et al., 2004; Huang et al., 2011; Navarrete & Fessler, 2006; Schaller & Neuberg, 2012). The standard explanation is that individuals from ethnic outgroups may harbor pathogens that individuals from the ingroup have not developed immunity against, so cues such as skin color and other physical features associated with such ethnic outgroups may elicit psychological responses and behaviors that at least indirectly limit exposure to individuals from these groups (e.g., negative attitudes, prejudice, hostility, avoidance). van Leeuwen and Petersen (2018) called this the "adaptation-for-outgroups" account because it holds that the behavioral immune system was designed to process foreign appearance as an infection cue (i.e., a proper domain cue) given the costs of foreign pathogens. Alternatively, the behavioral immune system may include design features for learning and transmitting information about avoiding pathogens, which could include stereotypes and experiences linking ethnic outgroups with infection (Fessler et al., 2015). In this way, the behavioral immune system could incorporate associations between ethnic outgroup cues and pathogens in order to mitigate infection risks (real or perceived), even if the behavioral immune system is not designed specifically to motivate avoidance of such groups (van Leeuwen & Petersen, 2018).

In sum, a diverse body of research has been motivated by the hypothesis that the behavioral immune system evolved to motivate the avoidance of people who display correlates of infectious disease. Importantly, given uncertainty inherent in detecting infection from imperfect cues and the asymmetric costs of mistakes, the behavioral immune system responds to

true infection indicators (proper domain cues) as well as cues that share perceptual properties of such indicators (actual domain cues). Such a sensitive system may help people effectively avoid infectious others, but it may also underlie unfounded prejudices toward people bearing anomalous yet objectively benign physical features (e.g., port-wine stain birthmarks, obesity, foreign appearance). That is, over-perception can be functional yet problematic.

My dissertation work expands on this over-perception hypothesis. The assumptions and empirical research reviewed in chapter one serve as a foundation for chapters two and three. In chapter two, I elaborate on why I should expect the behavioral immune system to treat some infection cues as stronger indicators of infection risk than others. If one cue better diagnoses infection risk or perhaps better diagnoses a more fitness-costly infection risk than another cue (e.g., a deadly or debilitating infection), then I should expect the behavioral immune system to react more strongly to that cue and any cue that shares many perceptual features with the strongly diagnostic cue. In three studies, I test this hypothesis by comparing the strength of association between infection concepts and obesity to the association between infection concepts and facial disfigurement.

In chapter three, I elaborate on why I should expect people to mentally represent infected others and I employ three mental representation methods to explore two hypotheses about how they might do that: the *threat-specificity hypothesis* and the *threat-combination hypothesis*. The threat-specificity hypothesis posits that people mentally represent infected others only with features that diagnose infection risk. Based on the behavioral immune system literature, these features include but are not limited to germiness, disfigurement, obesity, old-age, and foreignness. Importantly, based on the threat-specificity hypothesis, other threatening features like anger, dominance, violence, and muscularity should not appear in mental representations of

infected others. In contrast, the threat-combination hypothesis posits that people mentally represent infected others with a combination of threatening features, such as disfigurement as well as anger. Together, chapters two and three shed light on basic and unexplored mechanisms of the behavioral immune system.

Chapter II. Differences in Infection Cue Strength

As I discussed in more detail in chapter one, researchers who study pathogen avoidance psychology assume the behavioral immune system is a functional mental module: it comprises cognitive mechanisms with specialized functions and restricted inputs. This means that, for the behavioral immune system, pathogen threat-specific cues elicit psychological responses and behaviors whose functions serve to reduce pathogen threats. Pathogen threat-specific cues comprise true indicators of infection—the "proper" domain of the behavioral immune system—as well as cues that share perceptual properties with those true indicators. The set of proper domain cues that the behavioral immune system was designed to process plus the cues the system is able to process as a by-product of its design make up the "actual" domain of the behavioral immune system. Theoretically, all cues that make up the actual domain of the behavioral immune system should elicit psychological responses and behaviors that function to reduce pathogen threats.

A growing body of evidence suggests that a variety of anomalous cues—whether or not they truly indicate infection—elicit psychological response and behaviors that serve to reduce pathogen threats. That is, cues like rashes, runny noses, facial discoloration, wrinkles, crossed-eyes, swelling, obesity, and physical disability—among other anomalous cues—elicit visual attention, disgust, prejudice, and avoidance, psychologies and behaviors which can all play a role in reducing infection risk. Though these findings provide evidence for an effect of anomalous cues on behavioral immune responses, it is at least unclear from current theory whether the

behavioral immune system should treat these cues and other benign, anomalous cues as equally indicative of infection risk. One might expect this system to discriminate between anomalous, benign facial features (e.g., more looking, disgust, avoidance) if some benign features closely resemble highly diagnostic symptoms more than others. On the one hand, if some features are highly diagnostic—they strongly distinguish between infected and non-infected individuals—then those features warrant uninhibited behavioral immune responses. For example, facial rashes are associated with a large variety of fitness reducing infections (e.g., Drage, 1999). On the other hand, if some features are poorly diagnostic, the behavioral immune system could mute its responses to them so people can explore fitness enhancing social interactions with people presenting with such features. For example, obesity has been linked to some infections and might weaken the efficacy of vaccines (Coetzee et al., 2009; Falagas & Kompoti, 2006; Mancuso, 2013; Whigham et al., 2006), but many more people meet criteria for obesity but harbor no infection. Moreover, obesity is a relatively modern physical anomaly that may not be under strong selection pressure (Pontzer et al., 2012), making it susceptible to over perception from the behavioral immune system. Such a poorly diagnostic infection cue must be interpreted in the context of other cues that diagnose beneficial social opportunities.

If rashes diagnose infection but obesity does not, then rashes fall within the proper domain of the behavioral immune system and obesity does not, and people might react with stronger behavioral immune responses to proper domain cues like rashes than to actual (but not proper) domain cues like obesity possibly because the behavioral immune system can more easily process inputs it was designed to process (Barrett, 2012; Barrett & Kurzban, 2006). Moreover, if benign facial disfigurements strongly resemble rashes, then, even though facial disfigurements do not fall within the proper domain of the behavioral immune system, people

may nonetheless react with stronger behavioral immune responses to benign facial disfigurements than to obesity. I test this hypothesis in four studies.

Study 1

My goal for Study 1 was to test whether people more strongly associate infectious concepts with facial disfigurement than with obesity. To measure associations, participants completed one of two versions of a custom Implicit Association Test—I call it an Anomaly-Infection IAT—where they categorized infectious concepts (e.g., epidemic) or harmless concepts (e.g., typewriter) in combination with faces bearing or not bearing one of two anomalous facial features: disfigurement or obesity. This implicit approach mimics that used in prior behavioral immune literature (e.g., Park et al., 2007) and allows us to estimate the strength of various associations while minimizing participant response biases.

Method

Participants. Between September 27 and November 22, 2017, 283 undergraduate students from the University of Michigan psychology subject pool participated in my study. I report pertinent sample information for Study 1 and all other studies in Table 1. I excluded participants from my final sample if they did not complete the full set of measures, made errors on more than 50% of IAT trials, or completed more than 10% of IAT trials in less than 300 milliseconds (Greenwald et al., 2003; Wolsiefer et al., 2017).

Table 1. Participant and stimulus characteristics for Studies 1-4.

Study	Participants	Age	Women	White
1	283	18.90 (2.16)	167 (59%)	171 (60%)
2	256	18.90 (1.11)	152 (59%)	165 (65%)
3 & 4	134	18.84 (0.90)	87 (65%)	99 (74%)

Note. The table displays study number, total number of participants, mean age (standard deviation in parentheses), number of (percentage) excluded participants, number of (percentage) women, number of (percentage) white participants, and total number of stimuli.

Procedure. Following consent, participants were randomly assigned to read instructions for one of two versions of the Anomaly-Infection IAT: an Obesity-Infection IAT or a Disfigurement-Infection IAT. As with a standard Implicit Associations Test, participants categorized words and images according to pairs of category labels as quickly and accurately as possible. Specifically, participants pressed one of two computer keys to classify words as either Harmless or Infectious, and, depending on their version of the IAT, to classify faces as (1) Average or Obese or as (2) Average or Disfigured. Depending on the type of trial, the keys to categorize faces shared the keys to categorize words. For example, on some trials, participants used the same key to categorize Harmless words and Disfigured faces. To ensure understanding, participants read definitions of Harmless and Infectious word categories prior to beginning the test.

Harmless Definition: Not harmful or injurious; not likely to irritate or offend; benign or innocuous.

Infectious Definition: Causing or communicating infection; tending to spread (disease) from one to another; contagious.

The Anomaly-Infection IAT featured blocks of 28 single-category learning trials, 28 paired-category practice trials, and 56 test trials; otherwise, my IAT procedure followed the standard design found in Greenwald et al. (2003) (see Table 2). For each participant, practice and test trials began with the category pairs of Harmless- Anomaly vs. Infectious- Average or pairs of Infectious-Anomaly vs. Harmless-Average. Following this, category pairs were swapped and participants completed additional blocks of practice and test trials. Because the standard IAT effect depends on the order in which category pairs are seen, this order was randomized between participants. Finally, participants completed measures of individual differences (see below), demographic questions, and open-ended questions about their experience in the study.

Table 2. Sequence of Trial Blocks in our Anomaly-Infection IAT (Study 1).

Block	Trials	Function	Left Key	Right Key
1	28	Learning	Harmless Words	Infectious Words
2	28	Learning	Anomalous Faces	Average Faces
3	28	Practice	Harmless Words + Anomalous Faces	Infectious Words + Average Faces
4	56	Test	Harmless Words + Anomalous Faces	Infectious Words + Average Faces
5	28	Learning	Average Faces	Anomalous Faces
6	28	Practice	Harmless Words + Average Faces	Infectious Words + Anomalous Faces
7	56	Test	Harmless Words + Average Faces	Infectious Words + Anomalous Faces

Notes. Function indicates the purpose of the trials in a given block. Left and Right Key columns indicate which response was paired with the left or right keyboard button. Anomalous faces were either Obese faces or Disfigured faces, depending on condition.

Face Stimuli. I selected 8 white men's faces from the Chicago Face Database (Ma et al., 2015). I used men's faces in order to reduce the number of factors in my analyses and thereby increase power to detect an effect. I recruited an artist from www.fiverr.com to digitally manipulate each face twice to appear obese or to possess a port-wine stain birthmark (16 total images). I refer to these facial cues as Anomalous cues. I recruited 200 participants from

Amazon Mechanical Turk to rate all 9 faces under all face-type conditions (Disfigured, Obese, and No Manipulation) on a list of subjective features: Disfigured, Germy, Fat, Heavy, Unpleasant, Upset, Happy, Clear, Realistic, and Photoshopped. As expected, the participants rated the Obese faces as significantly fatter and heavier (on average) than the Disfigured and Non-Manipulated faces, and they rated the Disfigured faces as significantly more disfigured (on average) than the Obese and Non-Manipulated faces (see Figure 1). Unexpectedly, participants rated the Obese faces as significantly happier, more unpleasant, and more photoshopped than the Disfigured faces. Throughout my paper, I report analyses that do not control for these potential confounding variables, and I note whether conclusions change when I do control.

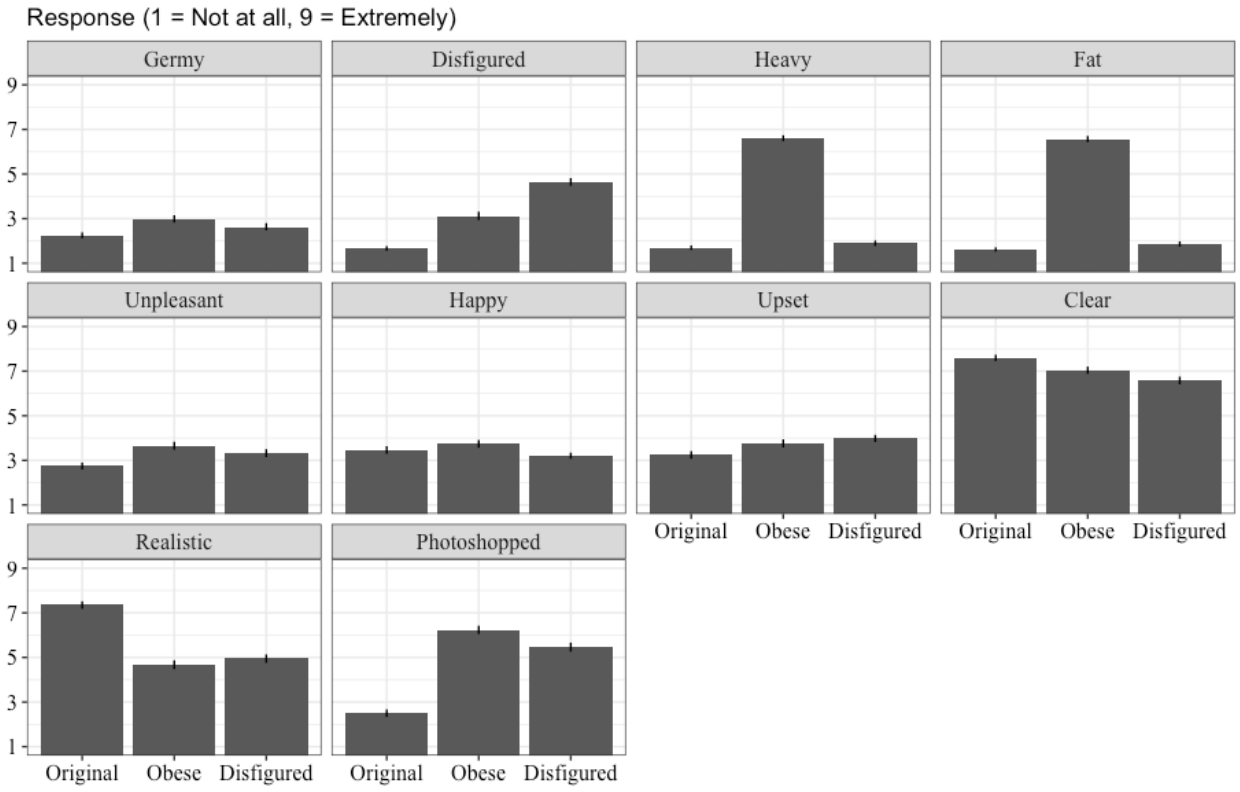

Figure 1. Pretest ratings by face type and rating variable. Bars represent average rating, collapsing over stimulus, and error bars represent bootstrapped 95% confidence intervals (1,000 resamples).

Word Stimuli. I used a variety of dictionaries and my own judgment to select 8 words that refer to Harmless objects or concepts and 8 words that refer to Infectious objects or concepts. The set of Harmless words served as a non-threatening comparison set. I conducted a post-hoc survey asking undergraduate participants to report whether they associate these words with the categories I assigned them to. Participants explicitly associated harmless words less with infection than with infection words. These words were used as stimuli in Studies 1-3.

Harmless Words: Tape, Cardboard, Typewriter, Dryer, Bran, Teaspoon, Sod, and Ground

Infectious Words: Epidemic, Germs, Plague, Contagion, Flu, Virus, Parasite, and Bacteria

Individual Differences Measures. I had participants complete individual differences scales that I could use to evaluate exploratory moderation hypotheses (e.g., automatic associations depend on trait infection concern). To assess individual differences in trait infection concern, I had participants complete the Perceived Vulnerability to Disease Questionnaire (Duncan et al., 2009) and the Three Domains of Disgust Scale (Tybur et al., 2009). To assess individual differences in attitudes about obese people, I had participants complete the Anti-Fat Attitudes Questionnaire (Crandall, 1994) as well as Height and Weight (which I used to calculate the Body Mass Index). Finally, I also assessed Negative Emotionality from the Big Five Inventory-2 (Soto & John, 2017).

Removed Trials. I removed 1280/46536 target trials (3%) (i.e., trials from Blocks 3, 4, 6, and 7) responded to in less than 400 milliseconds or responded to after 10,000 milliseconds (Greenwald et al., 2003; Martin, 2016).

Statistical Power. To compute statistical power, I used Power ANalysis for GEneral Anova (PANGEA) version 0.02 (Westfall, 2016), which provides analytical solutions (i.e., solved via equations rather than simulation) for user-specified contrasts. I specified a fixed factor for Condition (2 Levels: Obese or Disfigured), Congruent Category (2 Levels: Incongruent or Congruent), a fixed factor for Stimulus Category (4 Levels: Infectious Word (n = 8), Harmless Word (n = 8), Average Face (n = 8), Anomalous Face (n = 8), a random factor for Participant (n = 256), and a random factor for Stimulus (n = 8 per Stimulus Category). I further specified Stimulus as nested within Stimulus Category and Participant nested within Condition (Word or Face). I used the default values for variance estimates: var(error) + var(Participant * Stimulus * Condition) = 0.217, var(Stimulus * Congruent Category * Condition) = 0.043, and var(Participant * Congruent Category) = 0.087. Given my specifications, I had 80% power to detect Cohen's $d = 0.24$.

Analysis Plan. Implicit Association Tests assess differences in reaction times to classify stimuli during incongruent trials (here, the Anomaly category paired with the Harmless category) compared to congruent trials (the Anomaly category paired with the Infectious category). In order to account for variability in reaction time associated with participant, stimulus, and other IAT-specific factors, I fit linear mixed effects models that resemble those described by Wolsiefer et al. (2017), who extended the conventional scoring algorithm developed by Greenwald et al., 2003 to a mixed models framework. Specifically, in all analyses, I used a model that included fixed effects terms for between-subjects experimental condition (Obese = -0.5, Disfigured = 0.5), whether the trial was congruent (Congruent = -0.5, Incongruent = 0.5), trial order (Incongruent Trials First = -0.5, Congruent Trials First = 0.50), stimulus type (Word = -0.5, Face = 0.5), word category (Harmless = -0.5, Infectious = 0.5, Face = 0), and face category (Average = -0.5, Obese

= 0.5, Disfigured = 0.5, Word = 0). In addition, to adjust for general processing speed, a potential confound, I averaged each participant's reaction times across their learning trials (i.e., classifying only words or only faces) and included this participant-level average as a covariate. Unlike the D-score typically calculated for IATs, this approach adjusts the effect of interest without giving more weight to participants with less variable reaction times and without sacrificing the interpretation that longer average reaction times during incongruent vs. congruent trials reflect stronger automatic associations between infectious concepts and anomalous cues resembling historical infection indicators.

My initial models also included random terms for participant and stimuli intercepts as well as for participant and stimuli congruent effect slopes. When models failed to converge or produced singular fits, I removed random terms following recommendations from Bates et al. (2015) (also see Barr et al., 2013). Ultimately, I built these models to estimate the congruency effect.

Results

I first tested associations between anomalous cues and infection concepts as a replication of prior research and then examined differences in association strength depending on the type of anomalous cue.

Do people automatically associate anomalous cues with infectious concepts? On average, participants took 314 milliseconds longer, 95% CI [287, 343], Cohen's d = 0.47 (for d computation, see Judd et al., 2017), to classify words and faces when the anomalous category—either Obese or Disfigured—shared a response key with the Harmless category compared to the Infectious category (see Figure 2). This suggests that participants more strongly associated anomalous faces with the Infectious category than with Harmless category, and they more

strongly associated Infectious words with the anomalous category than with the Average category.

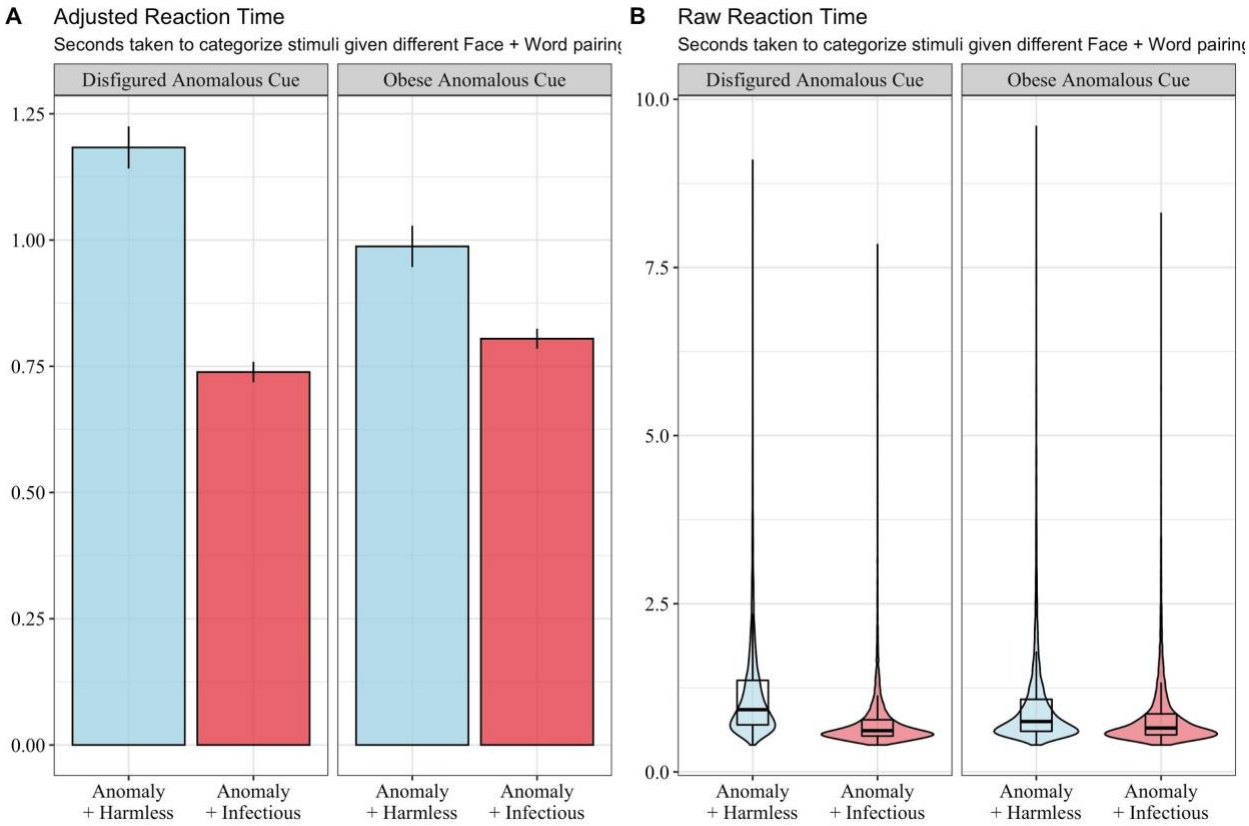

Figure 2. The bars in Panel A represent estimated reaction time means for incongruent (blue bars) and congruent trials (red bars) that have been adjusted for general processing speed, among other design factors. Bigger differences between bars (blue minus red) indicate stronger automatic associations for anomalous cues and infectious concepts. Error bars represent 95% confidence intervals based on the standard error of the Congruent x Anomalous Cue interaction (fit using the effects R package; (Fox & Hong, 2009; Fox & Weisberg, 2018). The mixture of violin and boxplots in Panel B represent the full range of raw reaction time within the different types of trials. The violin regions index the probability of observing reaction times in that region (i.e., wider regions mean scores are more probable there). Boxplots depict the median line, the interquartile range, and "whiskers" extending at most 1.5 times the interquartile range beyond the 25th and 75th percentiles (reaction times beyond that are considered extreme).

Does the automatic association between anomalous cues and infectious concepts depend on the type of anomalous cue? The congruency effect size depended on the type of anomalous cue, $M_{Difference}$ = 262 milliseconds, 95% CI [205, 321][1], Cohen's d = 0.39. On

[1] For main effects and interactions from the model, I report parametric bootstrapped 95% confidence intervals (1000 resamples; see confint.merMod from the lme4 R package (Bates et

average, participants took 445 milliseconds longer, 95% CI [405, 485], to classify words and faces when the Disfigured category shared a response key with the Harmless category compared to the Infectious category, whereas the corresponding difference in reaction time was 183 milliseconds, 95% CI [144, 223], when the anomalous category was Obese (see Figure 2). Thus, though participants automatically associated both Obese and Disfigured categories with Infectious concepts (and Average categories with Harmless concepts), this association was stronger between Disfigured categories and Infectious concepts.

Discussion

In Study 1, participants completed an Implicit Association Test in which they categorized Harmless and Infectious words and faces as either (1) Average or Obese or (2) Average or Disfigured. Obesity and Disfigurement served as anomalous cues that previous research suggests are associated with infection because these facial features deviate from human-typical morphology. Participants more strongly associated both types of anomalous cues with Infectious concepts than with Harmless concepts. But this difference in association strength also depended on the type of anomalous cue. This difference was larger for Disfigurement cues than for Obesity cues, supporting the hypothesis that people perceive facial disfigurement as a more diagnostic infection cue than obesity. However, Study 1 is limited because its design does not allow for an "uncontaminated" comparison between facial disfigurement to obesity. That is, due to the standard IAT design, the current findings could reflect stronger associations made between average faces and harmless concepts in the context of disfigured faces as compared to obese

al., 2014) but for conditional main effects (i.e., simple slopes), I report simultaneous confidence intervals based on the normal approximation.

faces. To disentangle this possibility, Study 2 used a modified IAT design in which participants could classify obesity and facial disfigurement cues within the same Implicit Association Test.

Study 2

My goal for Study 2 was, like Study 1, to test whether people more strongly associate infectious concepts with facial disfigurement than with obesity. But I designed my Study 2 Anomaly-Infection IAT to assess the association strength between both disfigurement and infection and obesity and infection in the same task. This allowed us to test my hypothesis within-subjects.

Method

Participants. Between February 2nd and March 29th, 2017, 256 undergraduate students from the University of Michigan psychology subject pool participated in my study. I report pertinent sample information for Study 2 and all other studies in Table 1. I used the same exclusions criteria from Study 1.

Procedure. The procedure for Study 2 resembled the procedure for Study 1 except I did not randomly assign participants to an Anomalous Cue condition. Instead, participants categorized faces as either Obese or Disfigured in the same Implicit Association Test. Thus, no Average category was used here.

Removed Trials. I removed 759/4308 target trials (18%) (i.e., trials from Blocks 3, 4, 6, and 7) responded to in less than 400 milliseconds or responded to after 10,000 milliseconds (Greenwald et al., 2003; Martin, 2016).

Statistical Power. Using PANGEA version 0.02 (Westfall, 2016), I specified a fixed factor for Condition (2 Levels: Incongruent or Congruent), a fixed factor for Stimulus Category

(4 Levels: Infectious Word (n = 8), Harmless Word (n = 8), Obese Face (n = 8), Disfigured Face (n = 8), a random factor for Participant (n = 256), and a random factor for Stimulus (n = 8 per Stimulus Category). I further specified Stimulus as nested within Stimulus Category (Word or Face). I used the default values for variance estimates: var(Error) = 0.20, var(Participant * Stimulus * Condition) = 0.05, var(Stimulus * Condition) = 0.10, and var(Participant * Condition) = 0.10. Given my specifications, I had 80% power to detect Cohen's *d* = 0.35.

Analysis Plan. Out analysis plan was similar to Study 1, except the anomalous cue contrast (Disfigured vs. Obese) was embedded in the congruency effect (Incongruent vs. Congruent): In half the trials, the Obese category shared a response key with the Infectious category (so the Disfigured category shared a response key with the Harmless category), and in the other half of the trials, the Disfigured category shared a response key with the Infectious category (so the Obese category shared a response key with the Harmless category). I coded the congruency contrast so that larger, more positive values mean that participants took longer when the Disfigured category shared a response key with the Harmless category than with the Infectious Category (Obese/Congruent = -0.5, Disfigured/Incongruent = 0.5).

Results

Do people more strongly associate infectious concepts with disfigurement or obesity? On average, participants took 141 milliseconds longer, 95% CI [119, 163], *d* = 0.25, to classify words and faces when the Disfigured category shared a response key with the Harmless category compared to the Infectious category (and when the Obese Category shared a response key with the Infectious category compared to the Harmless category) (see Figure 3). Thus, participants associated Infectious concepts more strongly with Disfigured faces than with Obese

faces (or participants associated Harmless concepts more strongly with Obese faces than with Disfigured faces).

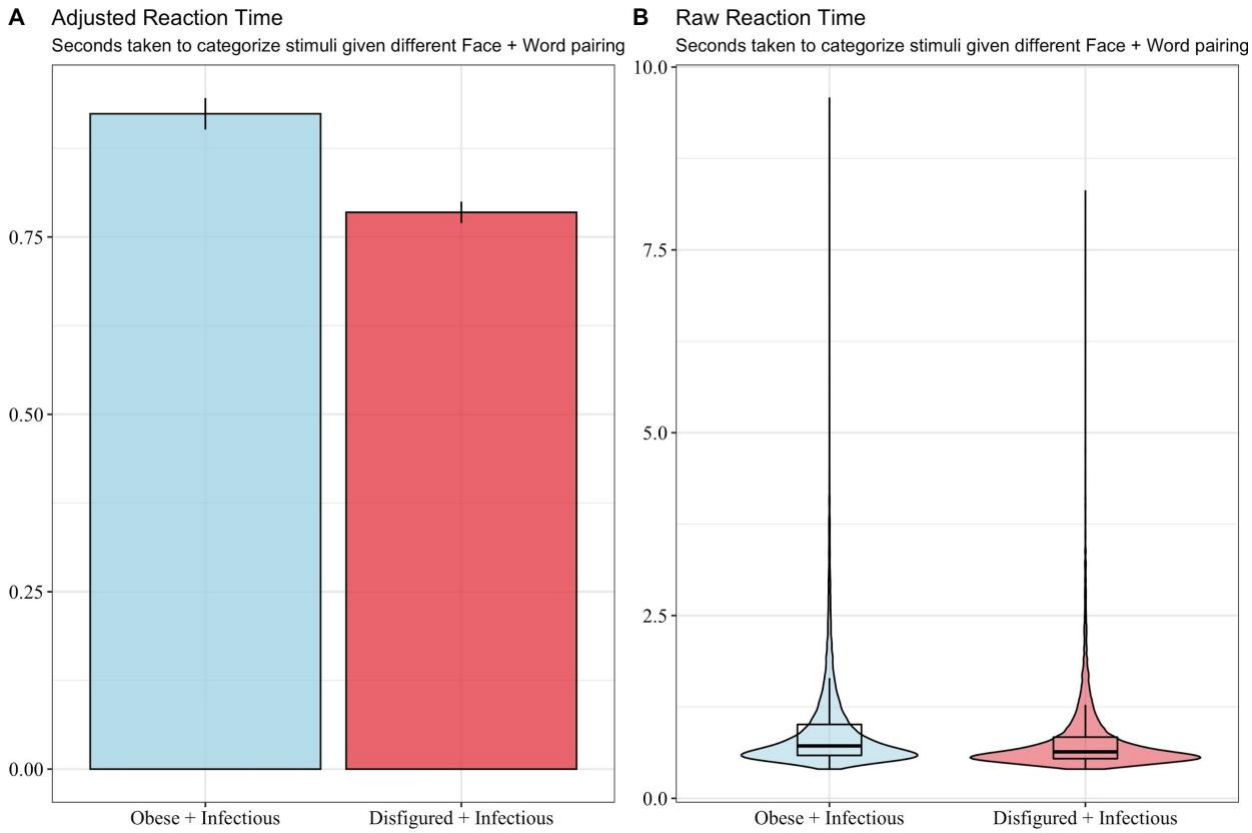

Figure 3. The bars in Panel A represent estimated reaction time means for incongruent (blue bars) and congruent trials (red bars) that have been adjusted for general processing speed, among other design factors. The difference in the height of the bars (blue minus red) indexes the automatic association between an anomalous cue and infectious concepts. Error bars represent 95% confidence intervals based on the standard error of the Congruent x Anomalous Cue interaction (fit using the effects R package; Fox & Hong, 2009; Fox & Weisberg, 2018). The mixture of violin and boxplots in Panel B represent the full range of raw reaction time within the different types of trials. The violin regions index the probability of observing reaction times in that region (i.e., wider regions mean scores are more probable there). Boxplots depict the median line, the interquartile range, and "whiskers" extending at most 1.5 times the interquartile range beyond the 25th and 75th percentiles (reaction times beyond that are considered extreme).

Discussion

As in Study 1, participants more strongly associated facial disfigurement with infectious concepts than they associated obesity with infectious concepts. Extending findings from Study 1, Study 2 showed this pattern when anomalous cue categories were directly compared. .

Two limitations are present in Studies 1 and 2, however. People may more strongly associate negative concepts with facial disfigurement than with obesity, regardless of whether those

negative concepts are infection-related or not. Moreover, I assumed that participants in the previous studies were aware that facial disfigurement is not infectious. It is possible that participants instead viewed the disfigurement cues as representing truly infectious hazards. To address these issues, in Studies 3A and 3B, I (1) made clear to participants that the facial disfigurement cues were benign and (2) tested whether another negative concept—laziness—is more strongly associated with disfigurement than obesity.

Studies 3A and 3B

Both Studies 3A and 3B were designed to address potential alternative interpretations for the findings in Studies 1 and 2. Participants in Study 3 were told the facial disfigurement cues they were tasked with classifying were benign burn scars, thereby making it clear that no anomalous cues were actually infectious. Study 3B replaced the Infectious category with a Lazy category, thereby providing a test of whether people more strongly associate any negative concept with disfigured faces compared to obese faces.

Method

Participants. Between March 12th and April 4th, 2018, 134 undergraduate students from the University of Michigan psychology subject pool participated in my study. I report pertinent sample information for Study 3 and all other studies in Table 1. I used the same exclusions criteria from Studies 1 and 2.

Procedure. Study procedures resembled the procedure in Study 2, except I randomly assigned participants to either categorize words as Harmless or Infectious and faces as Obese or Burn Scar (Study 3A), or categorize words as Harmless or Lazy and faces as Obese or

Disfigured (Study 3B). To assist categorizations in Study 3A, I defined Obese and Burn Scar for participants.

Obesity Definition: Obesity means having too much or excessive body fat, more than what's considered healthy for one's height. Obesity happens over time when one eats more calories than one uses.

Burn Scar Definition: Burns cause skin cells to die. Damaged skin produces a protein called collagen to repair itself. As the skin heals, thickened, discolored areas called scars form. Some scars are temporary and fade over time. Others are permanent.

Word Stimuli. For Study 3B, I used a variety of dictionaries and my own judgment to select 8 words that refer to Lazy objects. I also defined this term for participants.

Lazy Words: Procrastinating, Careless, Indifferent, Inactive, Sluggish, Neglectful, Passive, and Slacker

Lazy Definition: Unwilling to work or use energy; averse or disinclined to work, activity, or exertion; indolent.

Removed Trials. For Study 1A (Burn Scar vs. Obese), I removed 337/10416 target trials (3%) (i.e., trials from Blocks 3, 4, 6, and 7) responded to in less than 400 milliseconds or responded to after 10,000 milliseconds (Greenwald et al., 2003; Martin, 2016). For Study 1A (Lazy vs. Harmless), I removed 90/11760 target trials (1%) (i.e., trials from Blocks 3, 4, 6, and 7) responded to in less than 400 milliseconds or responded to after 10,000 milliseconds

Statistical power and analysis plan. I computed power and conducted analyses in Study 3 the same way as I did in Study 2. Given approximately 67 participants per condition and my other specifications, I had 80% power to detect Cohen's $d = 39$.

Results

Study 3A. Do people more strongly associate infectious concepts with benign burn scars or obesity? On average, participants in Study 3A took 173 milliseconds longer, 95% CI [137, 213], $d = 0.30$, to classify words and faces when the Burn Scar category shared a response key with the Harmless category compared to the Infectious category (and when the Obese Category shared a response key with the Infectious category compared to the Harmless category) (see Figure 4). Thus, participants associated Infectious concepts more strongly with benign Burn Scar faces than with Obese faces (or participants associated Harmless concepts more strongly with Obese faces than with benign Burn Scar faces).

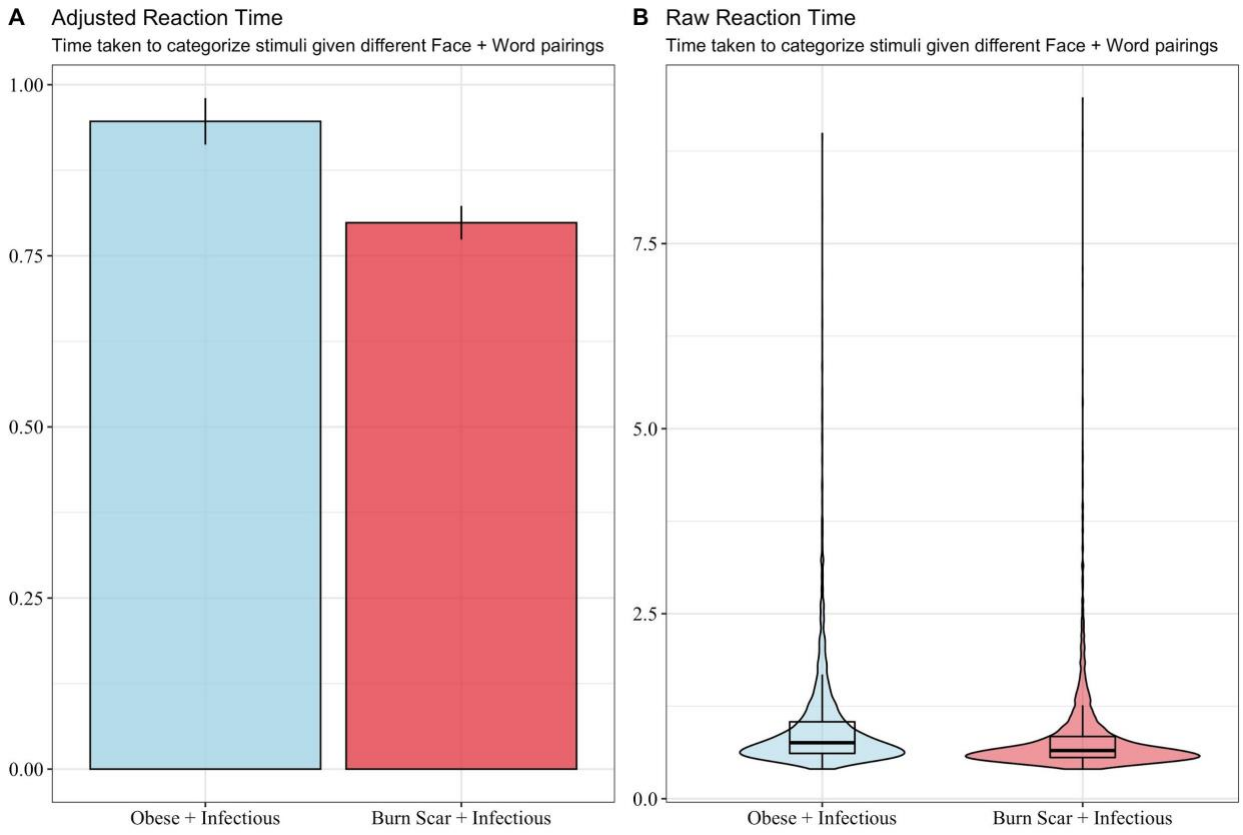

Figure 4. The bars in Panel A represent estimated reaction time means for incongruent (blue bars) and congruent trials (red bars) that have been adjusted for general processing speed, among other design factors (Study 3A). The difference in the height of the bars (blue minus red) indexes the automatic association between an anomalous cue and infectious concepts. Error bars represent 95% confidence intervals based on the standard error of the Congruent main effect (fit using the effects R package; Fox & Hong, 2009; Fox & Weisberg, 2018). The mixture of violin and boxplots in Panel B represent the full range of raw reaction time within the different types of trials. The violin regions index the probability of observing reaction times in that

region (i.e., wider regions mean scores are more probable there). Boxplots depict the median line, the interquartile range, and "whiskers" extending at most 1.5 times the interquartile range beyond the 25th and 75th percentiles (reaction times beyond that are considered extreme).

Study 3B. Do people more strongly associate lazy concepts with disfigured faces or obesity? On average, participants in Study 3B took 120 milliseconds longer, 95% CI [77, 163], *d* = 0.21, to classify words and faces when the Obese category shared a response key with the Lazy category compared to the Harmless category (and when the Disfigured Category shared a response key with the Harmless category compared to the Lazy category) (see Figure 5). Thus, participants associated Infectious concepts more strongly with benign Burn Scar faces than with Obese faces (or participants associated Harmless concepts more strongly with Obese faces than with benign Burn Scar faces).

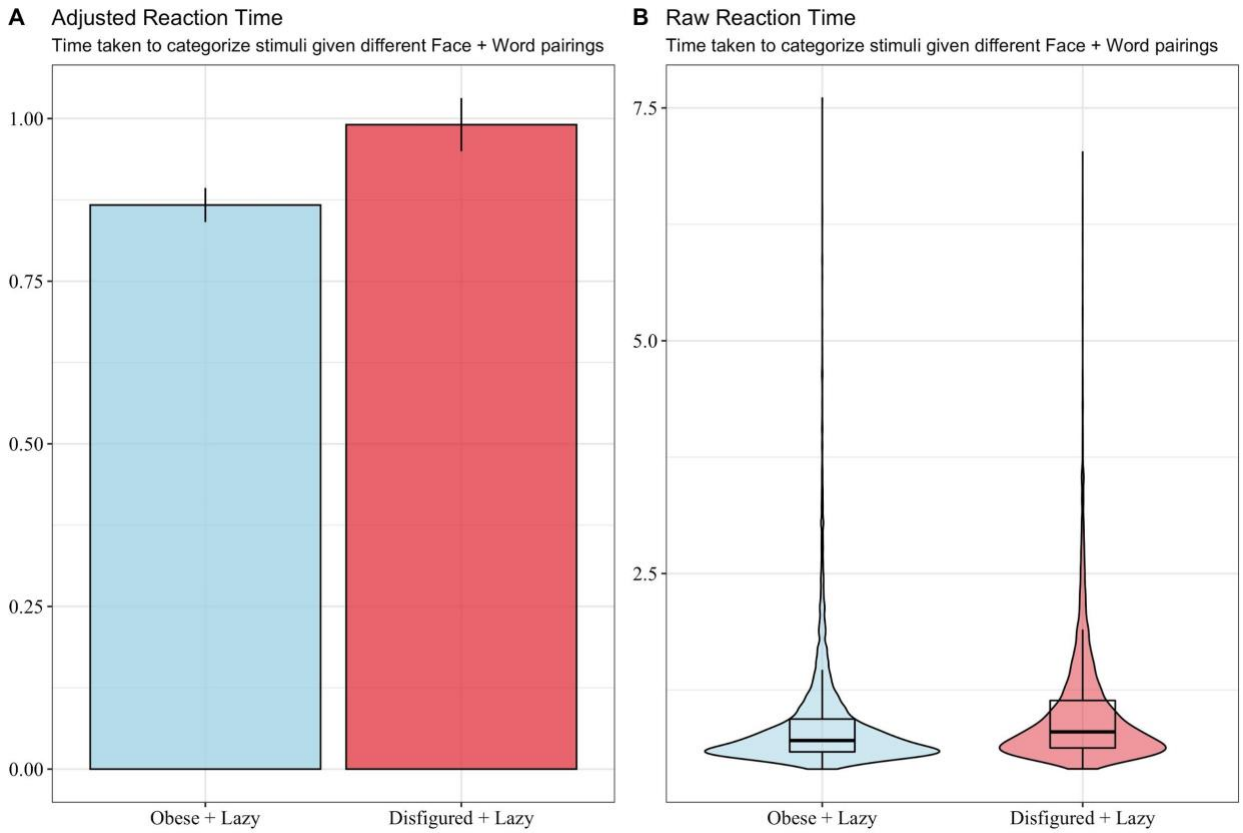

Figure 5. The bars in Panel A represent estimated reaction times means for incongruent (blue bars) and congruent trials (red bars) that have been adjusted for general processing speed, among other design factors (Study 3B). The difference in the height of the bars (blue minus red) indexes the automatic association between an anomalous cue and infectious concepts. Error bars

represent 95% confidence intervals based on the standard error of the Congruent main effect (fit using the effects R package; Fox & Hong, 2009; Fox & Weisberg, 2018). The mixture of violin and boxplots in Panel B represent the full range of raw reaction time within the different types of trials. The violin regions index the probability of observing reaction times in that region (i.e., wider regions mean scores are more probable there). Boxplots depict the median line, the interquartile range, and "whiskers" extending at most 1.5 times the interquartile range beyond the 25th and 75th percentiles (reaction times beyond that are considered extreme).

Discussion

Studies 3A and 3B helped to address alternative interpretations of my earlier findings. As in Studies 1 and 2, in Study 3A, participants more strongly associated infectious concepts with facial disfigurement than with obesity, even though participants were made aware that the facial disfigurements were caused by non-infectious burns. This result is inconsistent with the hypothesis that participants associate facial disfigurement with infectious concepts simply because they believe that the disfigurements were truly caused by infections. In addition, in Study 3B, participants more strongly associated lazy concepts with obesity than facial disfigurement. This result is inconsistent with the hypothesis that people more strongly associate any negative concepts with facial disfigurement than with obesity.

General Discussion

Across four studies, I investigated whether people more strongly associate infectious concepts with facial disfigurement than with obesity. Consistent with previous findings, I found that people associate infectious concepts with both facial disfigurement and obesity (Study 1). However, I found that this association is stronger for facial disfigurement compared to obesity (Studies 1-3). I found that the infection-disfigurement association was stronger than the infection-obesity association in a between-subjects experimental design (Study 1) as well as in a within-subjects design (Studies 2 and 3). In addition, I found the infection-disfigurement was stronger even when I told the participants the facial disfigurements represented benign (non-infectious) burn scars (Study 3A). Last, when I tested whether people simply more strongly

associate negative concepts with facial disfigurement than with obesity, I found that people more strongly associated lazy concepts with obesity than with facial disfigurement (Study 3B). Together, these findings provide preliminary evidence that people perceive some benign anomalous facial features as more diagnostic of infection than others.

Implications

These findings lend additional support to the pathogen detection error management hypothesis that people over-perceive infectious disease in anomalous cues (i.e., false positive errors) when infection indicators are noisy and false negative errors (i.e., deadly infections) are more fitness costly than false positives (e.g., missed social opportunities) (Haselton & Nettle, 2006; Nesse, 2005; Schaller, 2016). That is, the automatic associations I observed between infectious concepts and anomalous cues suggest my participants over-perceived infection in both obesity and facial disfigurement because infection is costly and at some level participants were uncertain whether these cues indicated the people presenting with them were infected. In addition, my finding that participants more strongly associated infection with disfigurement than with obesity is consistent with the uncertainty condition of the error management hypothesis. If participants were equally unsure whether disfigurement or obesity indicated infection, then I would predict equal behavioral immune system responses to these cues based on the error management hypothesis. But if one anomalous cue better diagnoses infection—or at least perceivers think it does—then I would predict that cue to elicit a stronger behavioral immune response because uncertainty has been reduced. So, my participants probably associated infection more strongly with facial disfigurement than with obesity because they were more certain facial disfigurement indicates infection (or that obesity does not indicate infection).

My findings also speak to which cues might make up the "proper" domain of the behavioral immune system (Schaller, 2016; Sperber, 1994; Sperber & Hirschfeld, 2004)(Schaller, 2016; Sperber, 1994; Sperber & Hirschfeld, 2004). If the behavioral immune system was designed to process cues truly indicative of infection (the proper domain of the functional module), then cues that more closely resemble those cues should be easier to process and thus should elicit a strong behavioral immune system response (e.g., staring, disgust, avoidance). Correspondingly, cues that share perceptual properties of proper domain cues but that nonetheless meaningfully deviate in appearance from truly diagnostic infection cues (the "actual" domain) should be harder to process and thus should elicit a weak behavioral immune response. This implies that a benign facial rash should elicit a practically identical behavioral immune response to a truly infectious facial rash if both rashes look the same. In contrast, if obesity only minimally resembles swelling caused by an infection or if obesity is not (and has not) strongly correlated with infection, then obesity might elicit a weaker behavioral immune response compared to more diagnostic cues. My finding that participants more strongly associated infection with facial disfigurement than with obesity is consistent with the hypothesis that facial disfigurement more closely resembles the behavioral immune system's proper domain inputs than obesity.

The possibility that the behavioral immune system responds more strongly to proper domain cues has practical implications. When researchers select infection cue stimuli for their studies, they might consider how closely their stimuli hew to theoretically proper domain inputs. Stimuli that more closely resemble proper domain inputs could elicit stronger behavioral immune effects and thus require fewer participants to detect statistically with desired power. Relatedly, if researchers are interested in more generalizable effects of the behavioral immune system, they

might include stimuli that vary how closely the resemble theoretically proper domain cues (Michalak & Ackerman, 2020). Importantly, cues that less closely resemble proper domain cues might elicit stronger effects for people higher in trait-level pathogen threat concern (i.e., interaction effects), and cues that more strongly resemble proper domain cues may elicit less variable effects across levels of trait threat concern (i.e., main effects) (Ackerman et al., 2018; Tybur et al., 2014). Last (and more speculatively), researchers interested in applying behavioral immune system hypotheses to stigma intervention research might consider the extent to which pathogen concerns explain variability in stigma directed toward people who present with anomalous facial features. Recall from my findings that people associated infectious concepts with obesity, but they more strongly associated obesity with lazy concepts. If a stigmatized facial feature does not closely resemble an infection cue, then pathogen avoidance psychology may play a smaller role in the stigma and might only weakly inform an intervention study design.

Limitations

My findings were limited in at least two important ways. First, my participant samples comprised only Michigan undergraduates (see Table 1), even my relatively large face stimuli sample comprised only young, white men, and I only investigated two anomalous facial features: obesity and disfigurement. Harmful pathogens and their symptoms present in the face are numerous and highly variable, and many benign features resemble these infection symptoms to variable degrees. Importantly, these facial features and how people perceive them vary across the world. My participant and stimuli samples by no means represent this rich global variation. Second, my dependent measures are limited to automatic associations between infectious concepts and anomalous facial features. Future research might investigate whether a variety of anomalous facial cues elicit similar patterns in other behavioral immune responses like attention,

disgust, and avoidance (e.g., Ackerman et al., 2009; Lieberman et al., 2012; Miller & Maner, 2012; Ryan et al., 2012). That said, my findings are relatively strong (i.e., four studies with large test statistics, small p-values) and thus lend preliminary support to the error management hypotheses I described.

Conclusion

Across four studies, I found evidence that people automatically associate infectious concepts with two anomalous facial features: facial disfigurement and obesity. However, people more strongly associated infectious concepts with facial disfigurement than with obesity. These findings lend initial support to the hypothesis that cues that more strongly resemble proper domain input to the behavioral immune system (i.e., highly diagnostic infection cues) elicit stronger behavioral immune response.

Chapter III. Mental Representations of Infected Others

What does an infected person look like? In your mind, you might picture someone pale and weak, someone with a runny nose and watery eyes, or someone with facial rashes and sores. Whichever features appeared in your mind's eye, where did they come from? If you are a health professional or researcher, you probably chose features that align with specific theories and hypotheses in your field. In contrast, if you are a layperson, you probably chose features based on intuition, experience, and stereotypes. Put differently, mental representations depend on expectations. These expectations constrain the breadth of the features we imagine, and, consequently, the features we study, as well as the methods and measurement tools we use to study them.

I propose that the choices researchers make based on their expectations, even when emerging from theory, can limit the ability of studies to wholly capture how people mentally represent aspects of others, such as threats. Moreover, even methods that enable participants to report how they mentally represent social categories (thereby minimizing the influence of researcher expectations) still reflect participant expectations. Thus, such methods may not match what those individuals spontaneously represent in their mind's eye. Do methods that allow for strong influences from expectations produce similar or different representations than methods that restrict the influence of expectations? I examine this question in the domain of pathogen threat psychology, where perceivers represent the faces of infected others.

I begin by reviewing how a functional perspective on threat management explains why perceivers orient to particular cues of threat (i.e., threat-specificity), and then I detail the

strengths and limitations of expectation-driven versus data-driven methods of threat assessment. To preview my empirical findings, when participants could easily apply stereotypic beliefs, their representations showed more threat-specificity than when such expectations were constrained by a data-driven reverse correlation task. These results suggest that our current understanding of threat management psychology may be limited by approaches that privilege expectations—of laypeople or of researchers—for choosing experimental stimuli and testing aspects of threat processing.

Functional Threat Management

People process and react to sick individuals differently than they do violent individuals. This is, in part, because effectively avoiding infection requires different behaviors than avoiding violence. For example, one washes their hands to avoid getting sick when interacting with someone who coughs and sneezes, whereas one raises their hands to avoid injury when interacting with someone who brandishes a weapon. Distinct threats entail distinct psychological and behavioral solutions. From what I refer to as the functional threat management perspective, natural selection has favored mental systems that enable people to perceive, feel, think, and behave in ways to reduce threats in particular rather than threats in general (Barrett, 2012; Cosmides & Tooby, 1994; Holbrook & Fessler, 2015; Neuberg et al., 2011; Tooby & Cosmides, 1992). This perspective does not imply that specific threats always elicit specific responses (e.g., most threats elicit anxiety due to shared processing mechanisms), but a fully general response to all threats would be less efficient than responses targeting the unique affordances of each threat.

Consistent with this perspective, a growing body of evidence suggests people exhibit functional responses to specific threat cues. For example, people expressed more disgust in response to, and were less willing to touch, objects that had been touched by people with visible

cues of influenza compared to the same objects that had been touched by visibly healthy targets (Ryan et al., 2012). People also expressed more disgust by and less willingness to touch those objects when the target people bore non-infectious facial blotches, suggesting participants perceived features that merely resembled infection cues as if they were true indicators; that perception made them avoid touching objects that had been "contaminated" through physical contact. Sensitivity to many such facial features, including disfigurement, discoloration, swelling, and wrinkles, has been connected to the experience of pathogen threat (Ackerman et al., 2018; Ryan et al., 2012). In studies examining threat from aggression, responses differ. For example, people estimated greater state and trait anger in men holding household items that could be used as weapons (e.g., garden sheers) compared with men holding objects that are less plausible as weapons (e.g., a watering can) (Holbrook et al., 2014). Thus, a threat-specific cue elicited a functional response: People perceived men holding plausible weapons as more prone to anger. Other research has linked aggression threat to formidable physical features such as size and weight (Fessler et al., 2012). Based on both theory and findings such as these, researchers have made the case that mental systems connected to disease avoidance and violence avoidance involve distinct emotional responses, cognitive associations, and neurobiology (Neuberg et al., 2011; Oaten et al., 2009; Schaller et al., 2003).

Limitations of Common Threat Management Methods

Although perspectives on threat management such as the functional perspective have generated rich and productive literatures, they have also motivated the use of research designs that face two inferential challenges for answering how people mentally represent threats. The first challenge emerges from factors researchers omit from their study designs. Creating study methods, measures, and stimuli based on researcher-expectations may obscure evidence for

effects and processes not associated with those expectations. For instance, pathogen threat researchers have used some combination of theory and intuition to select cues to investigate, including rashes, swelling, and lesions as well as physical anomalies that resemble such cues, like port-wine stain birthmarks, crossed-eyes, obesity, and wrinkled skin (Ackerman et al., 2009; Duncan et al., 2009; Park et al., 2007; Ryan et al., 2012). In a typical study, researchers examine a threat-specific cue by manipulating its presence, manipulating the motivational state of the perceiver, or by measuring evaluations of the cue and the perceiver. Researchers then assess attention, explicit and implicit attitudes, emotional responses, or other types of reactions. Results of such studies inform whether people react to the chosen cues or manipulations, but those results may not generalize to other cues and manipulations. This may not seem like much of a problem—research has to start somewhere. But if such results do not generalize to unmeasured yet threat-relevant cues, then claims about threat-specificity—a key inference made from the functional perspective—would be unknowingly limited to the findings of reported study designs. In other words, conclusions would be biased (to a degree) by researcher design choices. Consider an example from a different literature. Over a decade of research found support for the hypothesis that people use perceptions of warmth and competence to understand social groups (Fiske et al., 2002). But this research was limited by the ratings scales (e.g., friendly, smart) and social groups (e.g., Blacks, women) researchers used in their studies. When researchers gave participants the opportunity to spontaneously generate social groups and evaluate them using their own psychological dimensions, they found that participants organized a wide variety of social groups using two novel dimensions: low-high socioeconomic status and conservative-progressive beliefs (Koch et al., 2016). Relying primarily on researcher-driven design choices led to mistaken, or at least limited, conclusions.

A second inferential challenge emerges when equating perceiver reactions with perceiver representations. Perceivers may react to threatening features of stimuli chosen by researchers but not spontaneously include those features in their threat representations. For example, if presented with photos of faces varying in age, perceivers may rate the younger faces as looking more trustworthy (Zebrowitz & Franklin Jr, 2014). However, if asked to list visible features of a trustworthy person, perceivers may not spontaneously list youth as a feature of a trustworthy appearance. As findings from studies measuring reactions to specific cues chosen by researchers accumulate, researchers may begin to treat these cues as though they collectively embody how people represent threatening others in their minds. However, this conclusion suffers from the fact that people may react to features that are not present in their mental representations, and they may mentally represent threats with features that researchers have not examined in reaction-based studies. To elaborate on the potential problems associated with this issue, I conceptualize mental representations next.

Mental representations of threat

Mental representations of threatening others characterize the internal prototype of a threat—how those threats are construed in the mind—and they include a set of key features. First, mental representations combine information across multiple processing levels, from lower-level perception to conceptual knowledge and higher-level cognitive states (Freeman & Ambady, 2011). Second, mental representations emerge dynamically in that people continuously construct their representations from these multiple information sources. Third, mental representations are complex combinations of information that can be "seen" in people's minds (Farah, 1988; Haxby et al., 2000; Kanwisher et al., 1997; Mechelli et al., 2004). For example, visual cues (e.g., white skin, frown), information regarding social categories (e.g., adult white male), behaviors (e.g., he

coughed), traits (e.g., he is withdrawn), and affective evaluations (e.g., bad, yuck) meld together into a sort of "mental mush" (Carlston & Smith, 1996, p. 184) that forms a mental representation (Freeman & Ambady, 2011; Sherman, 1996; Wyer, 2007). In sum, mental representations are built continuously from multiple sources of information and can be visualized in the mind.

Mental representations are infrequently studied in the threat literature relative to piecemeal processes such as perceiver reactions, evaluations, and associations. Presumably, they can be understood from a functional threat management perspective. To the extent that threats are processed in a threat-specific manner, representations should contain evidence of threat-specific features. Operationally, this specificity requires some features to be both associated with threat and distinctive of particular threats. Whereas a mental image of an infected person may appear sick and blotchy, a mental image of a violent person may appear angry and intimidating. To the extent that threats are processed more generally, however, these threat representations should share features with each other. A better understanding of these representations would provide insight into how threats are processed holistically as compared to studies that target reactions to individual cues. Indeed, it may be critical to investigate mental representations in order to address important limitations of the threat literature described earlier.

Current Research

To address these limitations, I used multiple approaches to estimate mental representations of two threat categories: infected persons and violent persons. I evaluate results from each approach according to two hypotheses. Our first hypothesis—what I label the threat-specificity hypothesis—follows from the functional threat management perspective and thus is consistent with the expectations of researchers using this perspective. The threat-specificity hypothesis predicts that threat representations will primarily include cues specific to and

diagnostic of that threat. The second hypothesis—what I label the threat-combination hypothesis—is the natural complement to the first. The threat-combination hypothesis predicts threat representations will include a combination of threat cues common across multiple types of threats. In sum, threat-specificity predicts representations will appear to pose one kind of threat (e.g., strong infection-related cues), whereas threat-combination predicts representations will appear to pose 2 or more kinds of threat (e.g., equally strong infection- and violence-related cues).

To test these hypotheses, I use two types of empirical approaches. First, in studies following an expectation-driven approach, participants listed traits they "saw" when imagining what infected and violent persons look like (Study 1) and drew infected and violent persons' faces (Study 2). These are common methods of assessing mental representations (Andersen & Klatzky, 1987; Stangor et al., 1992; Stangor & Lange, 1994). Qualitatively, these data gave us insight into which features come to mind when participants think of infected others in contrast to violent others. Much like researcher expectations can influence estimates of mental representations, participant expectations can, too. Participant-generated responses privilege the beliefs of perceivers in that perceivers are likely to deliberately edit their responses based on their own intuitions or stereotypes about what a given social category entails. To the extent that people expect threats to be distinct, substantial threat-specificity should characterize the resulting mental representations.

The second approach attempts to constrain perceiver (and researcher) expectations. For multiple reasons (e.g., insufficient access to internal representations), perceivers may report content that fits normative expectations but is not representative of how a social category appears in their mind's eye. To address this, I used a data-driven approach by leveraging reverse

correlation methods. Reverse correlation methods exploit the relationship between stimulus and response. Whereas more typical approaches estimate the correlation between fixed, researcher-selected stimulus attributes, and participant responses, "reverse" correlation approaches estimate the correlation between random stimulus attributes and participant selections (Brinkman et al., 2017; Dotsch & Todorov, 2012b; Mangini & Biederman, 2004; Todorov et al., 2011). For example, in the 2-image forced choice image classification task—a particular reverse correlation task—a base face is overlaid with random digital noise masks to represent many versions of that face with variable facial attributes. Perceivers choose from pairs of such faces the one that best represents the target category (e.g., Infected). Researchers then create a classification image by averaging the noise patterns from those perceiver choices and applying that average noise pattern to the original base image. This classification image serves as a visual proxy of the social category representation. Importantly, these images are participant-selected, visually compelling, and emerge from relatively more spontaneous mental processes. Classification images can then be rated along any number of dimensions to determine the features they possess. This approach has been used to estimate (visual) mental representations many unique categories, including racial and minimal out-groups (Dotsch et al., 2008; Ratner et al., 2014), welfare recipients (Brown-Iannuzzi et al., 2017), and atheists (Brown-Iannuzzi et al., 2018). Unlike more common approaches, the reverse correlation method has received comparatively less attention in the threat processing literature. I used this method in Studies 3-5 as a comparison to the expectation-driven approaches in earlier studies. Each of these studies includes two phases: In Phase 1, participants generated proxy mental images via the 2-image forced choice image classification task, and in Phase 2, independent participants rated features of those images generated in Phase 1.

Overall, I report five studies using a multi-method approach. I have ordered these to begin with studies utilizing methods that mimic researcher-driven expectations, where participants can freely apply their beliefs about social categories through deliberation, editing, extensive use of time, and so on. The second set of studies uses reverse correlation methods to constrain (though not entirely eliminate) these factors. For all but Phase 1 in Studies 2 and 3, I preregistered research questions, predictions, sampling plans, exclusion criteria, and analyses. I also report sensitivity analyses (i.e., compute the detectable effect size at 80% power given sample size and α) and provide empirical benchmarks for detectable effect sizes. For certain analyses, I deviated from our preregistered plans to reduce the number of reported tests and to synthesize measures in a conceptually meaningful way. In our supplemental repository (Michalak & Ackerman, 2017; https://osf.io/84vdp/), I also include a spreadsheet detailing original and revised plans (i.e., those reported in this article). I also make available preregistrations, materials, and de-identified data for all studies, including data sets not reported here. Finally, I include additional analyses in our repository (e.g., individual differences effects within each study).

Study 1

Our first goal in Study 1 was to give participants the opportunity to report which visible features they see in their mind's eye when they envision infected and violent others in order to test the threat-specific and threat-combination hypotheses. Here, participants simply listed visible traits they believed correspond with each of two threat categories. Our second goal was to evaluate whether the listed set of features supported the threat-specificity or threat-combination hypothesis. The threat-specificity hypothesis predicts participants would list only features

specific to and diagnostic of that threat, whereas the threat-combination hypothesis predicts participants would list a combination of features common across multiple types of threats.

Method

Listing visible traits of Germy and Violent persons

Participants. I recruited undergraduate psychology student participants between March 25th, 2019 and September 25th, 2019 (see Table 3 for pertinent sample characteristics).

Table 3. Reprinted from Michalak and Ackerman (2020). Characteristics of participant roles in Studies 1-5.

Study	Source	Role	N	Excluded	Age	Women	White
1	Undergraduate	Lister	117	0 (0%)	18.84 (0.81)	49 (42%)	64 (55%)
2	Undergraduate	Artist	147	0 (0%)	19.00 (0.90)	97 (66%)	90 (61%)
2	MTurk	Rater	129	14 (10%)	36.55 (11.74)	76 (59%)	99 (77%)
3	Undergraduate	Chooser	94	2 (2%)	19.13 (1.70)	45 (48%)	72 (77%)
3	MTurk	Rater	272	18 (6%)	35.68 (11.07)	139 (51%)	197 (72%)
4	Undergraduate	Chooser	205	0 (0%)	18.75 (0.92)	153 (75%)	180 (88%)
4	MTurk	Rater	464	50 (10%)	37.97 (12.27)	242 (52%)	301 (65%)
5	MTurk	Chooser	200	44 (18%)	36.94 (11.38)	92 (46%)	149 (75%)
5	MTurk	Rater	505	37 (7%)	35.46 (10.88)	272 (54%)	384 (76%)

Note. See supplement for additional information.

Statistical power. I are interested in whether some words are used more or less frequently to describe traits for Infected persons than for Violent persons (i.e., proportion differences). Power formulas for the difference between proportions require a non-intuitive arcsine transformation to obtain Cohen's h, which I report and explain next. Participants used a total of 3335 trait words. This many words affords 80% power to detect Cohen's h = 0.07 (pwr.2p.test function in the pwr package in R; Champely, 2018). Importantly, the size of Cohen's h depends on the difference in proportions and on the size of the proportions compared.

For example, a 2% difference from 50% corresponds to a larger Cohen's h (h = 0.04) than a 2% difference from 5% (h = 0.08).

Procedure. After consenting to participate, participants read:

"In this task, we'd like you to imagine two specific kinds of people. Specifically, what do these kinds of people look like? When you imagine either of these kinds of people, what traits can you see these people having?"

Next, participants listed in separate text boxes up to 10 visible traits for an Infected person and 10 visible traits for a Violent person, each on a separate page. At the top of each page, participants read, "What does an infected [violent] person look like?" above a definition of infected [violent]:

infected: affected or contaminated (a person, organ, wound, etc.) with disease-producing germs or pathogens; capable of causing infection in other people

violent: prone to commit acts of violence; uses physical force intended to hurt, damage, or kill someone or something

Results

Analysis Plan. I used the tidytext and wordcloud R packages (Fellows, 2018; Queiroz et al., 2019) to extract and count the number of words from participants' lists of traits (connecting i.e. stop words such as "in" and "with" were excluded). Next, research assistants categorized synonyms under single, common terms (e.g., angry and anger were both coded as angry). I then computed word frequencies by dividing the count of specific words within each threat category (e.g., weak, angry) by the count of all words used within that threat category. Finally, I tested differences in these proportions between threat categories using Fisher's Exact Test.

Top traits listed within threat categories. Participants most frequently listed eye, pale, tired, weak, sick, red, nose, skin, coughing, and sweaty as visible traits for Infected others (see Figure 6). The words eye, red, skin, and nose require context to interpret. Each involved qualifiers associated with sickness states. For example, eye included “drowsy eyed” and “heavy eyebags,” red included “red cheeks” and “red nose,” skin included “greenish/yellowish tinted skin” and “pale grey skin,” and nose included “runny nose” and “blowing nose.” In contrast, participants most frequently listed angry, face, eye, looking, strong, aggressive, muscular, dark, mean, and big as visible traits for Violent others. The words face, eye, looking, and dark also involved relevant qualifiers. For example, face included “angry face” and “scowling face,” eye included “scary, dark eyes” and “angry eyebrows,” looking included “angry looking” and “mean-looking,” and dark included “dark, scary eyes” or was listed by itself. Note that eye was used in both Infected and Violent responses, but for different reasons.

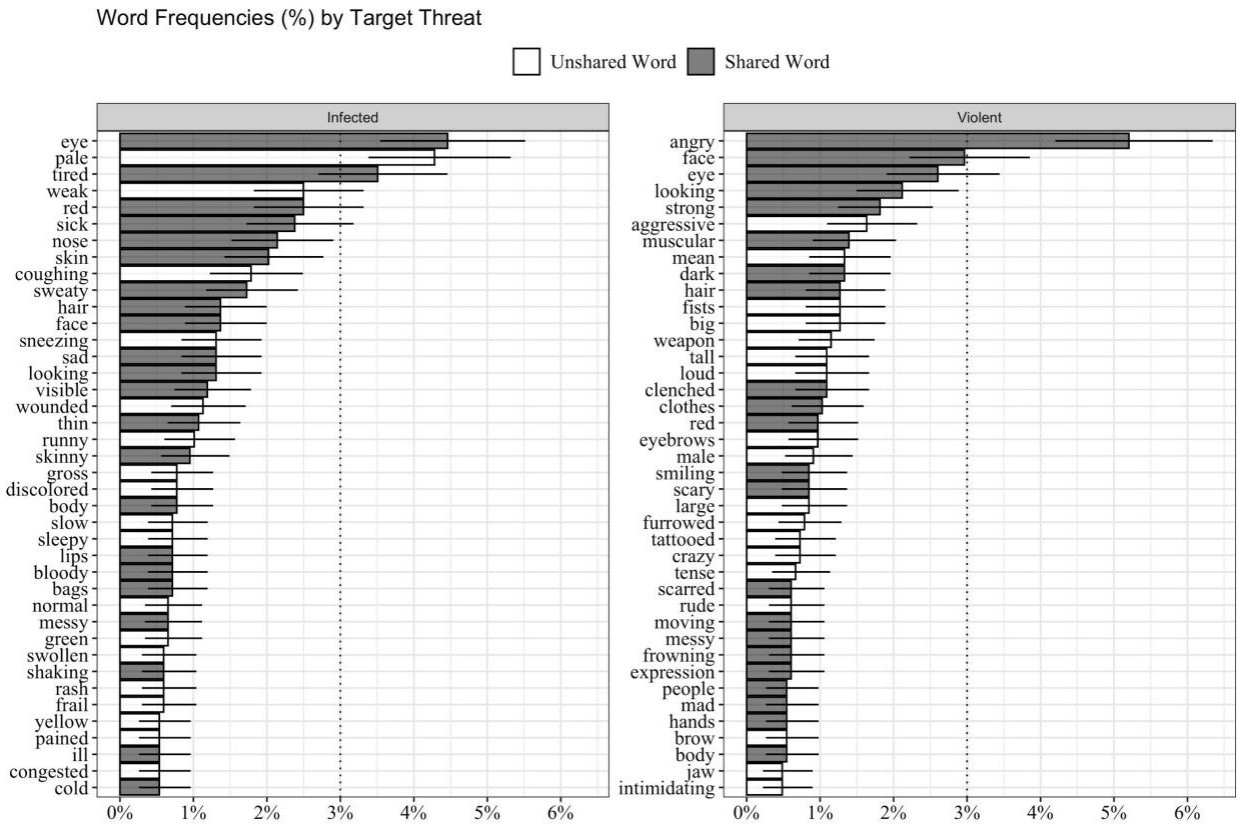

Figure 6. Reprinted from Michalak and Ackerman (2020). Each panel displays as proportions (x-axis) the top 40 most frequently listed visible traits (y-axis) for the Infected target (left panel) or Violent target (right panel). Bar fill indicates whether the word was shared (dark grey) or not (white) across threat categories (e.g., the word "eye" is a shared word because it was used to describe traits of both threat categories). Error bars represent 95% profile confidence limits. I added a dotted line at 3% for reference comparing across panels.

The most frequently used words in Infected trait responses were either not used in the Violent trait responses (pale and weak), used significantly less frequently in the Violent responses (tired and sick), or were used to describe different impressions than in the Violent responses (eye was used to describe "drowsy eyed" for an Infected trait but "scary, dark eyes" for a Violent trait) (see Table 4). Similarly, the most frequently used words in the Violent trait responses were either not used in the Infected trait responses (aggressive, mean), used significantly less frequently in the Infected responses (angry, strong, muscular), or were used to describe different impressions than in the Infected responses (face was used to describe

"scowling face" for a Violent trait but "pale face" for an Infected trait). Taken together, the trait listings suggest that people envision visible traits relatively specific to each threat category.

Table 4. Reprinted from Michalak and Ackerman (2020). Between-threat category comparisons among top 10 most frequently used words within each threat category (Study 1).

Word	Infected	Violent	*OR*	Lower	Upper	*h*
Eye	4.46%	2.60%	1.75	1.18	2.62	**0.10**
Pale	4.28%	0.00%	∞	19.47	∞	**0.42**
Tired	3.51%	0.06%	60.02	10.34	2384.35	**0.33**
Red	2.50%	0.97%	2.62	1.44	5.01	**0.12**
Weak	2.50%	0.00%	∞	10.95	∞	**0.32**
Sick	2.38%	0.12%	20.10	5.20	171.68	**0.24**
Nose	2.14%	0.12%	18.05	4.63	154.80	**0.22**
Skin	2.02%	0.18%	11.34	3.56	57.83	**0.20**
Coughing	1.78%	0.00%	∞	7.63	∞	**0.27**
Sweaty	1.72%	0.42%	4.12	1.76	11.18	**0.13**
Angry	0.18%	5.20%	0.03	0.01	0.10	**-0.38**
Face	1.37%	2.96%	0.45	0.26	0.76	**-0.11**
Looking	1.31%	2.12%	0.61	0.34	1.08	-0.06
Strong	0.06%	1.81%	0.03	0.00	0.19	**-0.22**
Aggressive	0.00%	1.63%	∞	∞	0.14	**-0.26**
Muscular	0.12%	1.39%	0.08	0.01	0.34	**-0.17**
Dark	0.36%	1.33%	0.27	0.09	0.68	**-0.11**
Mean	0.00%	1.33%	∞	∞	0.18	**-0.23**
Big	0.00%	1.27%	∞	∞	0.19	**-0.23**

Notes. Words are ordered by frequency within each threat category. The word "eye" topped both lists, so we table 19 instead of 20 most frequent words. Grey highlights indicate Infected person words. *OR* represents the odds ratio; lower and upper together represent the 95% confidence interval limits for the OR; and *h* represents Cohen's *h (***bolded** *h* values are significantly different at $\alpha = .05$). Last, ∞ represents infinity or undefined because one of the proportions in the odds ratio was exactly 0.

Discussion

Participants in Study 1 listed visible traits they expected infected and violent others to have. These form an expectation-driven representation of each category. Participants more often listed infection-related traits for infected others than for violent others, and they more often listed

violence-related traits for violent others than for infected others. Among the top words used to describe traits, words used in both threat categories were usually qualified in ways suggesting qualitatively different impressions. These results are most consistent with the threat-specificity hypothesis. Participants—like functional threat management researchers—expect infected others and violent others to possess visible traits that distinguish the kinds of threats those others pose.

Study 2

Study 2 is conceptually similar to Study 1 in that both studies allow participants to deliberate on the features to include in their mental representations. However, Study 2 allows participants to visually depict such representations. I asked participants to draw Germy (not Infected) and Violent persons in Phase 1, and then, in Phase 2, I had independent participants rate subjective features present in the drawings. These subjective ratings allowed us to test whether participants expect some features to be more strongly associated with infected people's appearance than other features (i.e., within-category comparisons), as well as whether some features better distinguish infected people's appearance from violent people's appearance (i.e., between-category comparisons). Finally, our last goal concerned the function of mental representations: I assessed whether people want to avoid the threatening people depicted in these mental images.

Method

Phase 1: Drawing faces of Germy and Violent persons

Participants. I recruited undergraduate psychology student participants between March 29th, 2017 and April 7th, 2017 (see Table 3 for pertinent sample characteristics).

Statistical power. I based power calculations in part on correlations between individual differences (see supplementary repository) and features to be judged later (in Phase 2). Casting these features as dependent measures, 147 artists afforded 80% power to detect Pearson's r = .23 (pwr.r.test function in the pwr package in R; Champely, 2018).

Procedure. I initially recruited participants for a study focusing on separate research questions about infectious disease (see the materials supplement in our repository). At the end of that study, participants completed the Perceived Vulnerability to Disease Questionnaire (Duncan et al., 2009) and saw a short debriefing page. Then I gave participants a pencil and a piece of paper with task instructions and a large oval on both sides. The instructions asked people to draw either a Germy face or a Violent face (manipulated within-subjects): "What does a germy [violent] person look like?

germy: full of germs; germ infested; appearing either sick or contaminated

violent: prone to commit acts of violence; uses physical force intended to hurt, damage, or kill someone or something

Please use the outline for a face below to draw a germy [violent] person."

I chose the Germy label after consulting with our undergraduate research assistants who believed that "Germy" would be most interpretable to undergraduate participants (in temporal order, this study was run prior to Study 1). See Figure 7 for example drawings (I make all our participants' drawings available in our online supplement). Among the 147 participants, 139 (94.56%) drew both a Germy and a Violent face. Our final stimuli sample comprised 139 pairs of drawings[2].

[2] By accident, 11 raters evaluated 1 or more blank drawings. We report results excluding these ratings. Mean differences are similar, and statistical significance decisions are the same whether or not we include these ratings data. See analysis supplement in our repository.

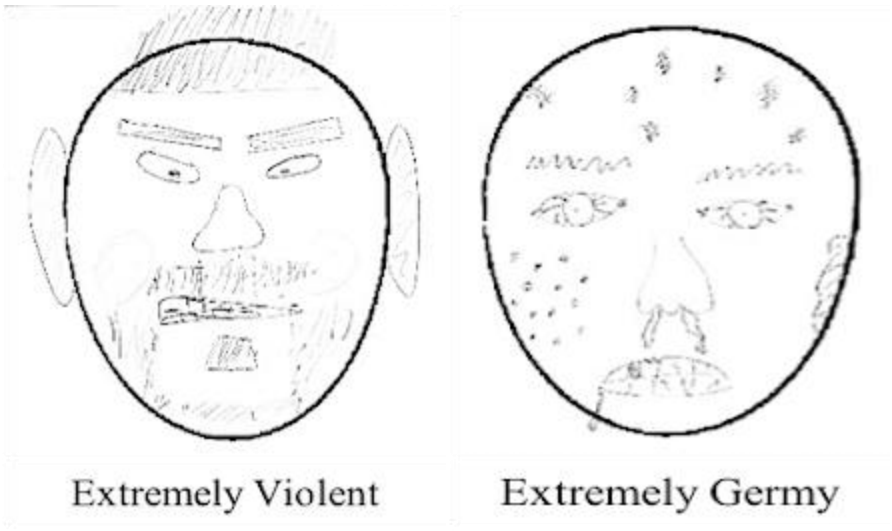

Figure 7. Reprinted from Michalak and Ackerman (2020). Images depict examples of a Germy drawing rated extremely Germy and a Violent drawing rated extremely violent (Study 2).

Phase 2: Measuring feature dimensions of drawn faces

Participants. I planned to recruit 138 MTurkers (assuming 10% would fail to meet inclusion criteria) using TurkPrime (Litman et al., 2017), leaving us with approximately 125 participants (see preregistration at https://osf.io/utyp6/). I sampled MTurkers until I had recruited approximately equal numbers of participants among 56 conditions and $N \geq 125$. One hundred forty-six participants opened our survey and I paid $0.50 to all 143 (97.95% completion rate) participants who submitted their MTurk HIT assignments (see Table 3 for pertinent sample characteristics).

Statistical power. Our final sample (N = 129) afforded us 80% power to detect Cohen's $d = 0.35$ for the Germy vs. Violent condition effect in the R(NCC) design described in Judd et al., 2017: Raters were nested within condition (they only saw Germy or Violent drawings) and artists were crossed with drawing category condition (artists drew both Germy and Violent faces); these settings combined to make multiple, unique sets of raters and artists. To compute this sample size value, I entered into the Shiny Web application (Westfall, 2016,

http://jakewestfall.org/two_factor_power/) that accompanies (Judd et al., 2017) the following values: total number of participants = 125, Total number of targets = 140, Total number of replications = 28, Residual Variance Partition Coefficient (VPC) = 0.4, Participant intercept VPC = 0.3, Stimulus intercept VPC = 0.2, and Stimulus slope VPC = 0.1.

Procedure. Research assistants digitally scanned drawings from Phase 1, cropped-out instructions, and adjusted image properties to increase visibility when needed. Because our stimuli set comprised 280 drawings, I divided this into 28 sets of five Germy drawings and 28 sets of five Violent drawings (i.e., 56 sets). Raters were randomly assigned to evaluate one of these sets (either Germy or Violent); so, raters saw one drawing per artist. Following consent, participants used a 9-point scale to rate each drawing on clarity (i.e., "How clear is this image?"), from 0 (Not at all) to 8 (Extremely). This was meant to familiarize participants with the drawings (Dotsch et al., 2008, Study 1). Next, participants completed the subjective feature dimension rating portion of the study. Participants read definitions for 12 feature dimensions they would use to rate the drawings: germy, disfigured, old, heavy, foreign, fatigued, healthy, violent, angry, dominant, muscular, and masculine. Participants were asked to confirm in a textbox at the bottom of the survey page that they read and understood the definitions. To reduce the number of items per screen, participants rated each drawing by itself on groups of four or five features at a time using the same 9-point scale they used for clarity ratings. Participants then used this scale to report their intentions to interact with the person represented in each classification image: "If you were to meet in real life, how much would you want to avoid physical contact with this person?" and "If you were to meet in real life, how willing would you be to stand near this person?" Finally, participants answered demographic questions, reported what they thought was the purpose of the study, and saw a short debriefing.

Results

Analysis plan. For our key analyses, I followed analyses for the R(NCC) design recommended by (Judd et al., 2017). For each rating, I fit a linear mixed effects model using the lmer function from the lme4 package in R (Bates et al., 2014); Satterthwaite degrees of freedom and p-values were calculated using the lmerTest package in R (Kuznetsova et al., 2017). Specifically, I regressed feature dimension rating (e.g., germy) onto Drawing Type (Violent = -0.5, Germy = 0.5), and I specified random intercepts for Rater, random intercepts for Artist, and random Drawing Type slopes for Artist (i.e., each Artist has their own Drawing Type effect).

Do Germy persons appear to have stronger infection-related features than less infection-related features? Comparing mean feature ratings within the Germy drawings, I found that the Germy drawings appeared to have stronger infection-related features than less infection-related features (top left plot of Figure 8; see supplement for pairwise tests). In particular, the Germy drawings received high average germy, fatigue, and unhealthy ratings (i.e., reverse-scored healthy).

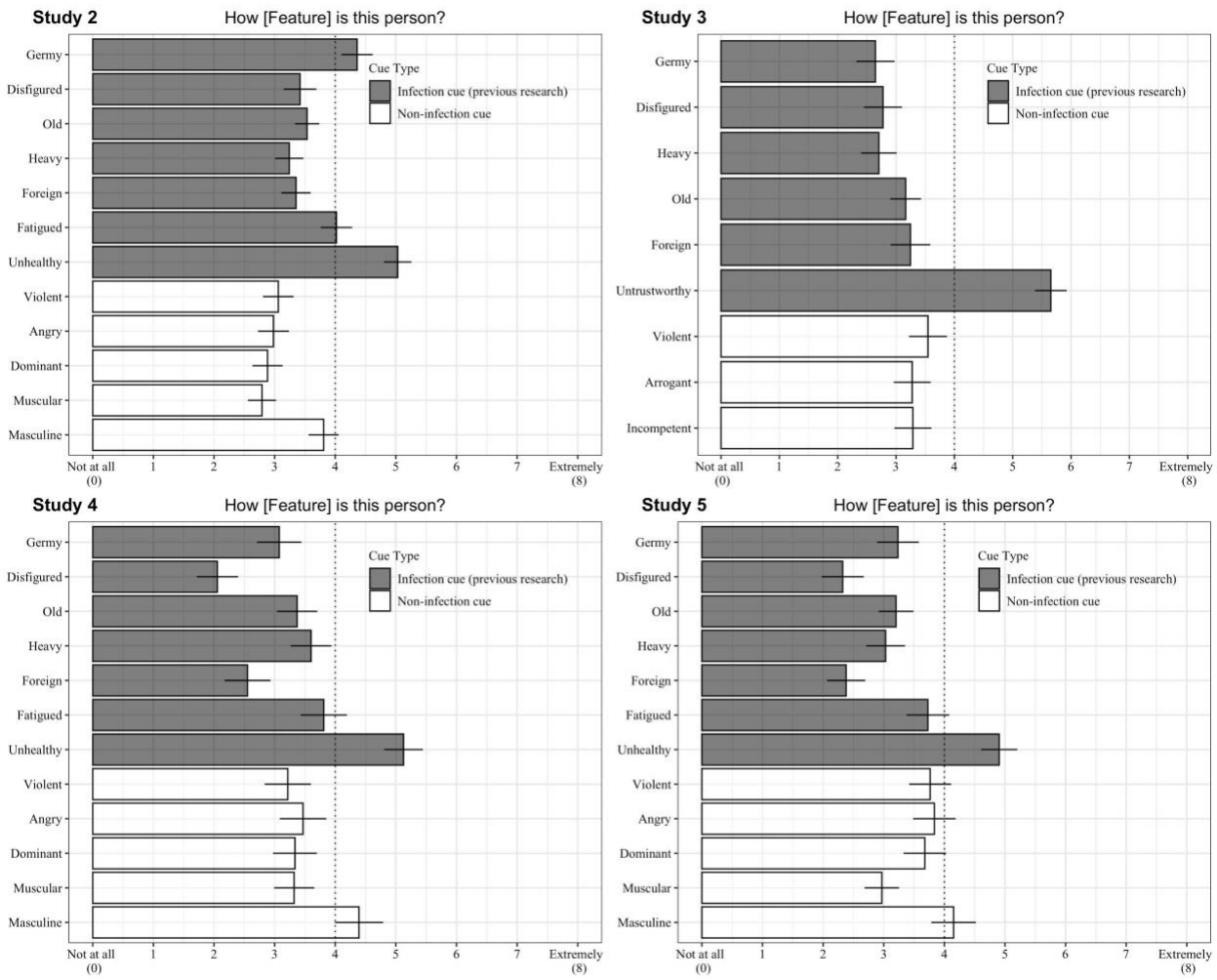

Figure 8. Reprinted from Michalak and Ackerman (2020). Mean feature ratings for the expectation-driven Germy representation (Study 2) and data-driven Germy and Infected representations (Studies 3-5). I colored the mean bars to highlight features that past research has categorized as cues associated with infection (grey fill) or has not examined in this context (white fill). I also reverse-scored trustworthy (untrustworthy) and healthy (unhealthy). The dotted line marks the middle of the response scale. Error bars represent 95% confidence intervals for individual feature means.

Do people draw Germy and Violent persons differently? Raters judged the Germy drawings to be significantly more germy (d = 0.32), less foreign (d = -0.27), less healthy (d = -0.40), less violent (d =-0.49), less angry (d = -0.64), less dominant (d = -0.62), less masculine (d = -0.39), and less muscular (d = -0.26) than the Violent drawings (see Figure 9 for mean differences and bootstrapped 95% confidence intervals; Bates et al., 2014). Additionally, raters judged the Germy drawings to be marginally less old (d = -0.20) but marginally more fatigued (d

= 0.18) than the Violent drawings. I found no sufficient evidence that raters judged the Germy drawings as more disfigured (d = -0.02) or heavier (d = -0.04) than the Violent drawings. In sum, I found evidence that people imagine Germy and Violent persons with many threat-specific features: They draw Germy persons with infectious disease cues—poor health and germiness—and they draw Violent persons with physical harm cues—foreignness, violence, anger, masculinity, and dominance. However, other features associated with pathogen avoidance responses (e.g., disfigured, heavy), did not significantly differ between the two sets of threat representations.

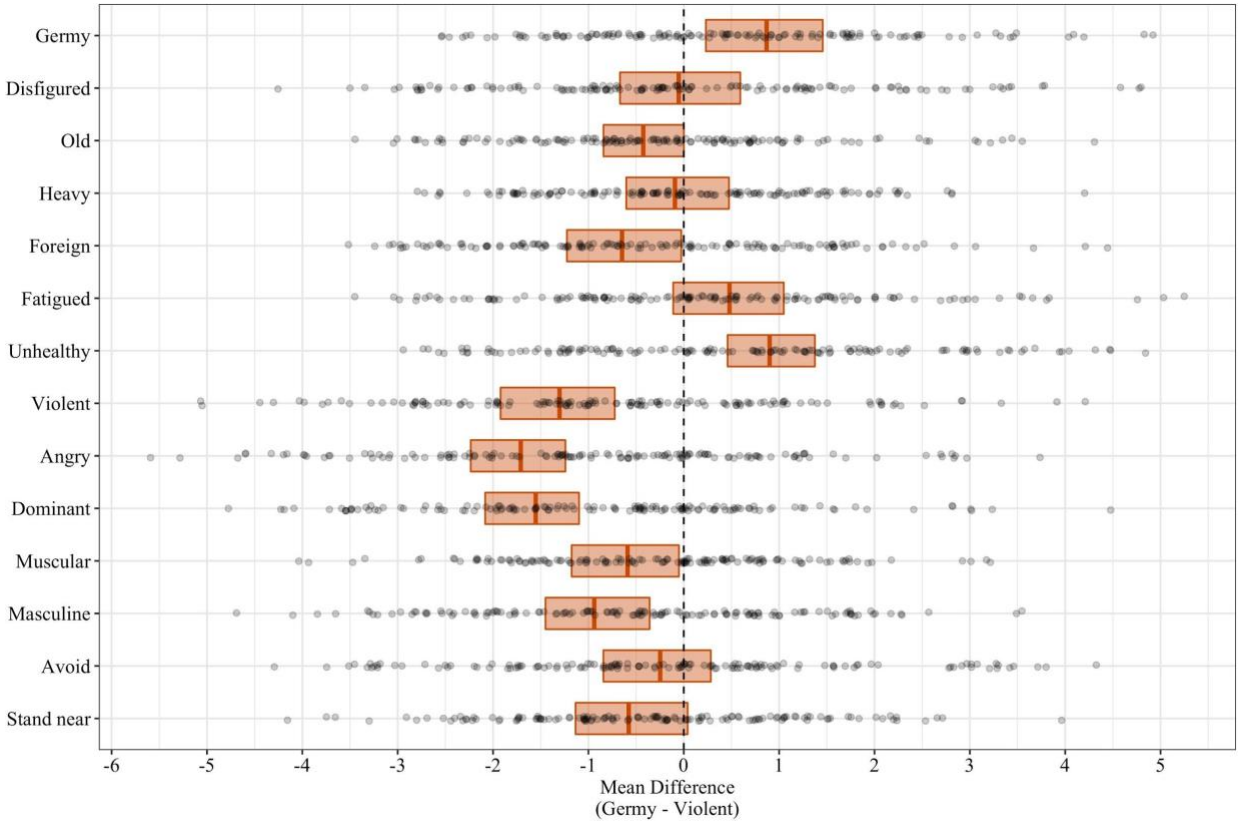

Figure 9. Reprinted from Michalak and Ackerman (2020). Mean differences in trait ratings between Germy and Violent drawings in Study 2. Dark, vertical lines inside crossbars (shaded boxes) depict fixed effects estimates for trait drawing condition (Germy – Violent). The widths of the crossbars represent 95% confidence intervals (bootstrap method, 1,000 resamples). Smaller, faded points depict observed difference scores (artist Germy drawing ratings minus their Violent drawing ratings). I also reverse-scored healthy (unhealthy).

Do people want to avoid Germy persons more than Violent persons? I found no sufficient evidence that raters wanted to avoid people in the Germy drawings more, $M_{Difference}$ 95% CI [-0.88, 0.37], than the people in the Violent drawings. However, raters were significantly less willing to stand near the people in the Germy drawings (d = -0.23).

Discussion

In Study 2, participants drew Germy and Violent people, representing their beliefs about what people who pose these threats look like. Examining within-category effects, Germy drawings appeared to have some stronger infection-related features (e.g., germy, disfigured, fatigued, unhealthy) than infection-unrelated features (e.g., violent, angry, dominant). For between-category effects, the Germy drawings in Study 2 were rated more germy, more fatigued, and less healthy than the Violent drawings (which were rated more violent, angry, old, and dominant than Germy drawings). This was not the case for other infection-related features previously examined in the pathogen avoidance literature (e.g., Germy drawings were rated less foreign and old than Violent drawings), though work in this literature does not typically compare cues between threat categories as I have done. Overall, data in the current study are most consistent with the threat-specificity hypothesis.

Taken together, results from the expectation-driven Studies 1 and 2 suggest participants represent Infected people with cues that are specific to and diagnostic of infection threat. In Study 1, participants listed visible traits for infected persons that are associated with infection and that distinguish an infected person from a violent person. In Study 2, participants drew infected persons with features that are associated with infection and that distinguish an infected person from a violent person. These findings are consistent with the threat-specificity hypothesis,

suggesting that representations derived from expectation-driven methods support hypotheses from the functional threat management perspective.

Study 3

To complement the expectation-driven approach in Studies 1 and 2, our first goal for Study 3 was to estimate people's mental images of an infected person while limiting the influence of participant (and researcher) expectations. Toward this goal, I recruited participants to complete a data-driven, 2-image forced choice reverse correlation task. In this task, participants selected hundreds of faces that they thought best represent a Germy person. I then averaged their selections to make a proxy Germy mental image. Unlike listing traits or drawing representations, the reverse correlation task does not allow participants to edit what their final representation looks like in order to bring it in line with their expectations. Thus, the influence of expectations is limited (though not entirely eliminated).

Our next goal was, as in Studies 1 and 2, to assess whether this data-driven representation appears primarily with features associated with infection and whether these features distinguish it from a non-infected person. Last, I assessed whether people want to avoid the person in the Germy mental image.

Method

Phase 1: Estimating mental images of a Germy person

Participants. I recruited undergraduate psychology student participants between February 26th, 2016 and March 31st, 2016 (see Table 3 for sample characteristics). I excluded 2 participants: One of these participants fell asleep during the main task and did not complete the task nor any questionnaires, and the other participant was legally blind.

Statistical power. Following our analysis rationale from Phase 1 from Study 2, 94 chooser participants afforded 80% power to detect Pearson's r = .28 (pwr.r.test function in the pwr package in R; Champely, 2018).

Procedure. For the 2-image forced choice reverse correlation image classification task, I first generated 400 pairs of stimuli using the rcicr package in R (Dotsch, 2016). Using the software, I generated each stimulus pair by superimposing a random visual noise mask (on one of the pair) and its negative (on the second of the pair) on our base image, a grey scale average of all male faces in the Karolinska Face Database (Lundqvist et al., 1998) (see Figure 10). For each trial of the task, participants saw in random order a pair of these superimposed images and were asked to choose the face that looked more "Germy," our target label representing an infected person. After completing the task, participants answered three questionnaires that assessed trait-level threat concerns (see supplemental repository).

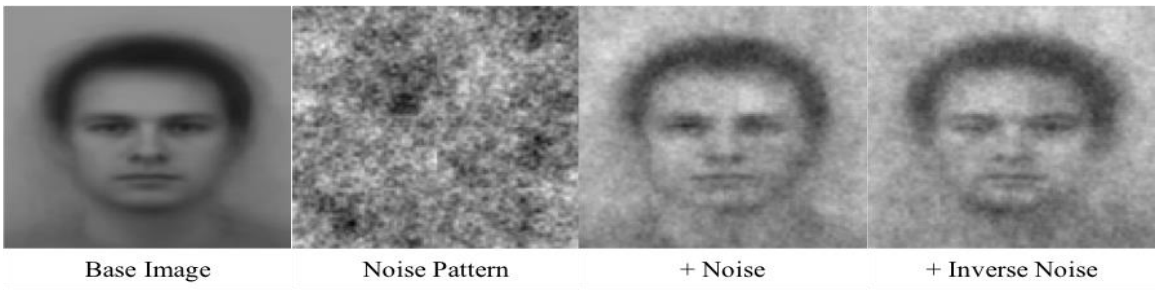

Figure 10. Reprinted from Michalak and Ackerman (2020). From left to right, the images depict our base face, a random noise pattern, an example noise pattern superimposed on the base face, and the inverse of the example noise pattern superimposed on the base face (Study 3).

Generating mental images: Classification images. Using rcicr (Dotsch, 2016), I created a classification image for this sample's mental image of a Germy person by first averaging the noise patterns of each participant's chosen faces, then averaging those averages across participants, and, finally, by superimposing that average onto our base face (see Figure 10). I did the same for the faces not selected—the anti-classification image, which represents this sample's

mental image of a non-Germy person. This distinction mimics comparisons described in previous research reports (e.g., welfare vs. non-welfare recipients; Brown-Iannuzzi et al., 2018). See the non-Germy and Germy classification images in Figure 11.

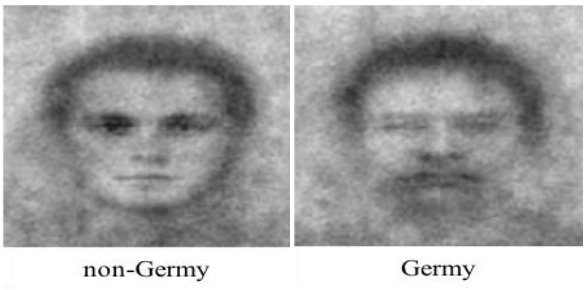

Figure 11. Reprinted from Michalak and Ackerman (2020). Images depict non-Germy and Germy classification images (Study 3).

Phase 2: Measuring subjective feature dimensions of Germy mental images

Participants. I planned to recruit at least 260 participants using TurkPrime (Litman et al., 2017) (see preregistration at https://aspredicted.org/xp2ey.pdf). Between December 28th, 2016 and December 29th, 2016, I sampled participants until I had recruited approximately equal numbers between conditions and $N \geq 260$. Participants opened our survey 350 times, and I paid $0.50 to all 290 who submitted their MTurk HIT assignments (82.86% submission rate). See Table 3 for pertinent sample characteristics.

Statistical power. Our final sample (N = 272) afforded us 80% power to detect Cohen's $d = 0.34$ between two independent samples (power.t.test function in the stats package in R; R Core Team, 2019), an effect size near the estimated median in intergroup processes research (Lovakov & Agadullina, 2017).

Procedure. I randomly assigned participants (the "raters") to rate either the Germy classification image (n = 135) or the anti-Germy classification image (n = 137). Participants also rated additional classification images created to test individual difference effects (see supplemental repository). Following consent, participants used a 9-point scale to rate each face

on clarity (i.e., “How clear is this image?”), from 0 (Not at all) to 8 (Extremely). This was meant to familiarize participants with the images (Dotsch et al., 2008, Study 1). Next, participants completed the subjective feature dimension rating portion of the study. Participants read definitions for nine feature dimensions they would use to rate the classification images: germy, disfigured, heavy, old, foreign, violent, arrogant, incompetent, and trustworthy. Participants were asked to confirm in a textbox at the bottom of the survey page that they read and understood the definitions. To reduce the number of items per screen, participants rated each classification image by itself on groups of four or five features at a time. Participants rated feature dimensions using the same 9-point scale they used for clarity ratings. Participants then used the same scale to report their intentions to interact with the person represented in each classification image: “If you were to meet in real life, how much would you want to avoid physical contact with this person?” and “If you were to meet in real life, how willing would you be to stand near this person?” In this way, participants first rated feature dimensions of each image, and then they reported their intentions to interact with the represented people. Finally, participants answered demographic questions, reported what they thought was the purpose of the study, and saw a short debriefing.

Results

Analysis plan. To test whether the Germy representation appeared to have stronger infection-related features than infection-unrelated features, I compared all pairwise feature mean ratings of only the Germy representation, correcting for multiple tests using a Bonferroni p-value adjustment (Maxwell, 1980).

To compare the non-Germy and Germy mental images on the variety of feature dimensions described above, I conducted a canonical discriminant analysis using the candisc function in the candisc package in R (Friendly & Fox, 2020), and I supplemented this analysis

with univariate analyses. Specifically, I used the discriminant procedure to compute a linear combination of weights that—when applied to our observed feature dimensions ratings—maximally distinguishes the non-Germy and Germy mental images. Conceptually, this allows us to distinguish the non-Germy and Germy mental images via a multivariate combination of feature dimensions—essentially a profile of features—rather than via each feature dimension by itself (i.e., ignoring correlations between features). Importantly, this procedure closely corresponds to a readily interpretable multivariate effect size, Mahalanobis's distance (D), which combines information from univariate effect sizes and correlations among the measures to index the standardized difference between two groups along the discriminant axis. Mahalanobis's D enjoys the same substantive interpretation as the widely used Cohen's d. Also, like Cohen's d, Mahalanobis's D can be converted to an overlap coefficient (e.g., $d = 0.85$ corresponds to 50% overlap between two univariate normal distributions). Researchers have used Mahalanobis's D to supplement standard univariate effect size analyses when testing gender differences in Big Five personality factors and facets (Del Giudice et al., 2012) as well as in implicit personality traits (Vianello et al., 2013). To address interpretation issues due to heterogeneity, I used a heterogeneity coefficient—the equivalent proportion of variables coefficient (EPV)—to help estimate whether only one or a few feature dimensions (i.e., small EPV coefficients) disproportionally account for observed multivariate differences between groups (Del Giudice, 2017).

Last, to test avoidance intentions (e.g., whether raters wanted to avoid the Germy representations more than the non-Germy representations), I conducted Welch's t-tests on the "want to avoid" and "willing to stand-near" items. In sum, I used a combination of canonical

discriminant analysis, univariate, and multivariate effect sizes to uncover the differences between non-Germy and Germy mental images.

Do Germy persons appear to have stronger infection-related features than less infection-related features? Comparing mean feature ratings within the Germy representation, I do not find sufficient evidence that the Germy representation appears to have stronger infection-related features than less infection-related features (top right plot of Figure 8; see supplement for pairwise tests).

Do people discriminate between non-Germy and Germy mental images? Using canonical discriminant analyses, I found that the nine feature dimensions combined well to maximally discriminate between the non-Germy and Germy mental images (cross-validated classification accuracy[3]: 78%, 95% CI [72%, 83%]), Canonical R2 = .28, $F(9, 270) = 11.57$, $p < .001$ (see Figure 12A). Along the single feature dimensions, raters judged the Germy mental image to be significantly more germy ($d = 0.39$), more disfigured ($d = 0.82$), heavier ($d = 0.63$), older ($d = 0.80$), more foreign ($d = 0.70$), more violent ($d = 0.31$), more incompetent ($d = 0.34$), and less trustworthy ($d = -0.51$) than the non-Germy mental image (see Figure 12B). Raters did not judge Germy representations to be significantly more arrogant than non-Germy representations ($d = -0.01$). When considering these feature dimension differences and their correlations together, I observed Mahalanobis's D = 1.25, 95% CI [0.94, 1.42], which corresponds to 36.35% overlap (assuming multivariate normality). In other words, subjective

3 We used a leave-one-out cross-validation procedure to fit a linear discriminant model (MASS R package, Venables & Ripley, 2002), and we used the posterior probabilities from that model (higher score = higher probability the participant rated a Germy representation) to calculate the area under the Receiver Operating Characteristic Curve (AUROC) (pROC R package, Robin et al., 2011), the "classification accuracy" value we report here and throughout this paper. Higher AUROC values indicate the model is better at distinguishing a Germy face from a non-Germy face. We computed confidence intervals around AUROC via a boostrapping procedure (2,000 stratified resamples).

perceptions of the non-Germy and Germy representations do not overlap much (less than 50%), suggesting that people do form relatively distinct Germy representations. Importantly, I observed some heterogeneity: The equivalent proportion of variables coefficient (EPV) suggests 51% (about 5) of the feature dimension ratings contribute equally to the multivariate effect, D. Specifically, it seems that perceptions of disfigurement, heaviness, age, and foreignness—the dimensions associated with the largest partial correlations—contributed most strongly to configural differences in non-Germy and Germy mental images.

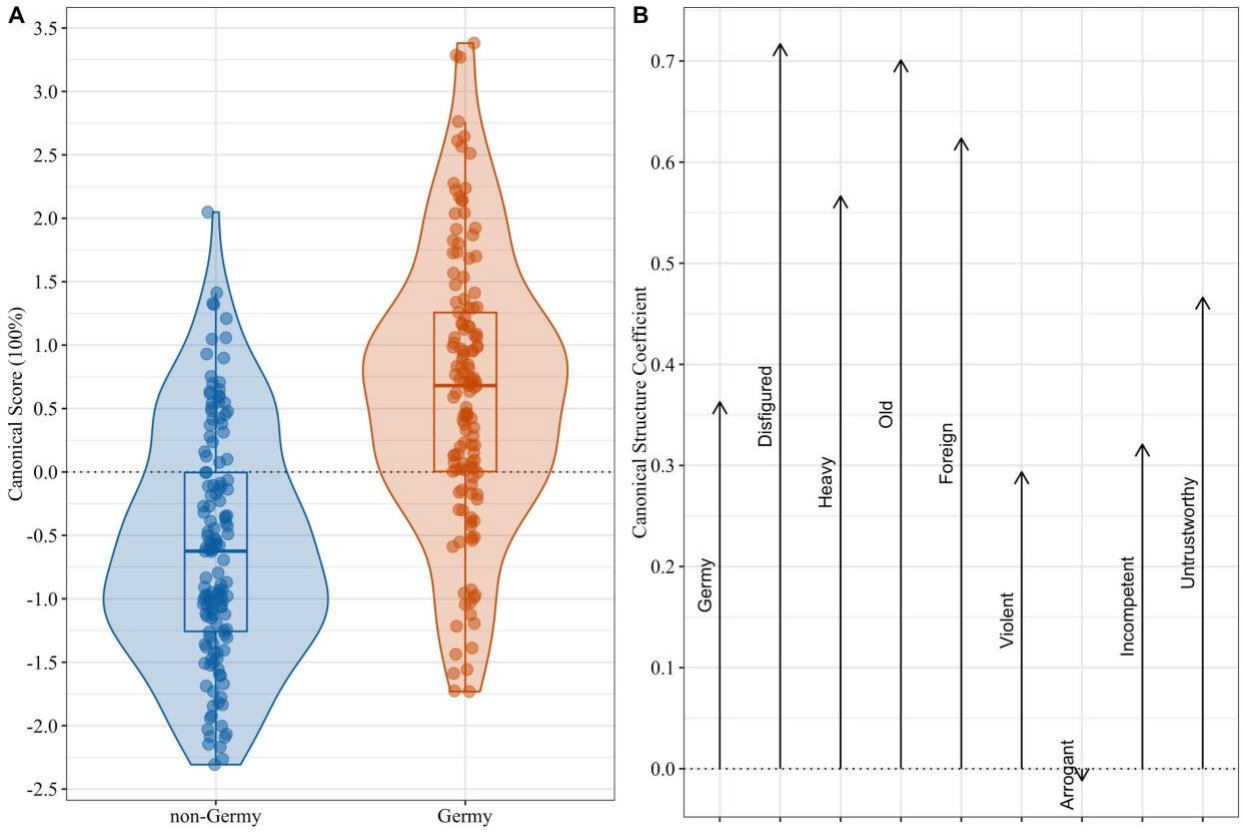

Figure 12. Reprinted from Michalak and Ackerman (2020). Together, the panels depict the maximal, configural difference between the non-Germy and Germy classification images (Panel A) and the relative contribution of each rating to that difference (Panel B) (Study 3). The canonical scores represent participant ratings transformed to maximize the difference in the canonical variable between the non-Germy and Germy conditions. Higher scores indicate a more Germy blend, and lower scores indicate a more non-Germy blend. Panel A depicts a combination of boxplots and violin plots that visualize representation canonical scores. Panel B depicts the direction and magnitude of partial (i.e., condition-adjusted) correlations between the individual feature dimension ratings and the canonical scores. Higher scores index the contribution of each feature dimension to the non-Germy/Germy differences.

Do people want to avoid Germy persons more than non-Germy persons? Raters reported wanting to avoid the person in the Germy mental image more, $M_{Difference}$ = 0.89, 95% CI [0.35, 1.44], t(276.53) = 3.21, p = .001, and they reported being willing to stand near the person in the Germy mental image less, $M_{Difference}$ = -0.51, 95% CI [-1.00, -0.02], t(278) = 2.07, p = .040, than the person in the non-Germy representation.

Discussion

In Study 3, participants generated mental images of Germy people through a reverse correlation image classification task. Importantly, the reverse correlation task is a data-driven method that limited participant's ability to edit their mental images in line with what they expect a Germy person to look like. Given this constraint on participants, do their estimated representations still appear threat-specific like those representations generated via expectation-driven methods used in Studies 1 and 2?

An independent sample of participants rated classification images on a variety of feature dimensions. Examining only the Germy representation, I did not find sufficient evidence that the Germy representation appeared to have stronger infection-related features than less infection-related features. However, when making multidimensional comparisons between face types, people did strongly distinguish between the Germy and non-Germy representations; disfigurement, heaviness, age, and foreignness ratings contributed most strongly to configural differences. The Germy representations also appeared more violent than the non-Germy representations, even though violent appearance is not a direct infection indicator. Last, the Germy representation appeared like someone people would want to avoid contact with, a motivation likely to reduce the threat posed by real infected people. Taken together, although the Germy representation appears to have threat-specific facial features that have been studied in the

pathogen avoidance literature, the Germy representation also appears to have a number of infection-unrelated negative features; thus, these ratings data are more consistent with the threat-combination hypothesis—a pattern different from that uncovered through expectation-driven methods in Studies 1 and 2.

Study 4

In Study 3, participants held a Germy mental image that—when compared to a non-Germy image—appeared to have many features associated with infectious disease but that also appeared violent, a trait that, at best, is only indirectly linked to infectious disease. One possibility is that these feature differences could simply be an artifact of comparing classification images to anti-classification images, which are mathematically opposite images (i.e., dark pixels in one image are light pixels in the other). However, this artifact explanation need not be true: Previous research on this reverse correlation image task suggests anti-classification images can be psychologically meaningful (e.g., submissive representations appear similar to anti-dominant representations, (Dotsch & Todorov, 2012a). Another possibility is that the label used to represent infection threat during the reverse correlation image classification task ("Germy") was imprecise, allowing participants to apply a variety of meanings during the image selection phase.

Given these possibilities, I made two key changes in Study 4. First, I replaced the Germy label with "Infected." Though similar to Germy, Infected unambiguously represents our social category of interest. Second, participants were assigned to choose faces representing either an Infected category or a Healthy category. By comparing representations for Infected and Healthy, rather than Infected and a composite of unchosen images (i.e., non-Infected), I can evaluate whether the differences found in Study 3 were merely procedural artifacts. If meaningful differences emerge between these categories, their pattern can be used to contrast the threat-

combination hypothesis and the threat-specific feature hypothesis. All other aspects of the study design followed those used in Study 3.

Method

Phase 1: Estimating mental images of Healthy and Infected persons

Participants. I recruited undergraduate psychology student participants between January 22nd, 2018 and March 12th, 2018 (see preregistration at https://osf.io/9t4pr). I tabled pertinent sample characteristics in Table 3. I randomly assigned these participants to either a Healthy (n = 103) or an Infected condition (n = 102).

Statistical power. The condition with the smallest sample size (n = 102) afforded 80% power to detect Pearson's r = .27 (pwr.r.test function in the pwr package in R; Champely, 2018).

Procedure. For the reverse correlation task, I first generated 400 new pairs of stimuli (i.e., new random noise patterns) using the rcicr package in R (Dotsch, 2016). I used the same base face and procedure I used in Study 3 with one design modification: Before starting the task, participants read instructions which included a definition of their randomly-assigned target category: (1) Healthy, "having or showing good health; not sick or injured; the condition of being well or free from infectious disease, " or (2) Infected, "affected or contaminated (a person, organ, wound, etc.) with disease-producing germs or pathogens; capable of causing infection in other people." Participants confirmed with a research assistant that they understood the definition. Then they completed a practice trial and confirmed again that they understood before starting the task. After completing the task, participants answered three trait-level threat concern questionnaires, answered demographic questions, and reported what they thought was the purpose of the study.

Generating classification images. Using rcicr as in Study 3, I created classification images and anti-classification images for participants' mental images of Infected and Healthy (see Figure 13). In total, I created one Infected classification image and one Healthy classification image as well as one Infected anti-classification image and one Healthy anti-classification image.

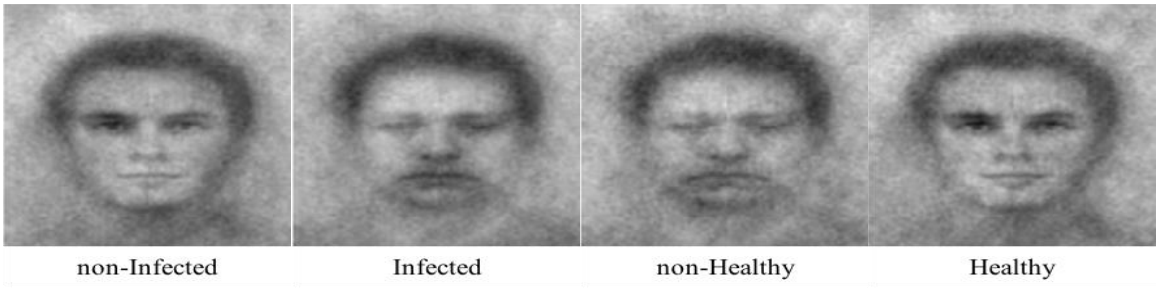

Figure 13. Reprinted from Michalak and Ackerman (2020). Images depict the non-Infected, Infected, non-Healthy, and Healthy classification images (Study 4).

Phase 2: Measuring features of Infected and Healthy mental images

Participants. I planned to recruit 444 MTurkers using TurkPrime (Litman et al., 2017) so I might collect a final N = 400 if 10% of participants failed to meet inclusion criteria (see preregistration at https://osf.io/uy4dm). Between March 24th, 2018 and March 25th, 2018, I sampled participants until I had recruited approximately equal numbers of MTurkers among four conditions and $N \geq 400$. Participants opened our survey 555 times and I paid $0.75 to all 464 (83.60% completion rate) participants who submitted their MTurk HIT assignments (see Table 3 for pertinent sample characteristics). I randomly assigned these participants to rate either the Healthy anti-classification image (n = 106), the Healthy classification image (n = 96), the Infected anti-classification image (n = 106), or the Infected classification image (n = 114). As in Study 3, participants also rated additional images assessing individual difference effects (see supplemental file).

Statistical power. Our power analyses reflect our preregistration in which I planned tests for all four conditions. Our final sample (N = 414) afforded us 80% power to detect Cohen's d = 0.27 for main effects (e.g., all Healthy conditions vs. all Infected conditions) and interactions and *d* = 0.39 for two-cell contrasts (power.t.test function in the stats package in R; R Core Team, 2019), effect sizes near the estimated median in intergroup processes research (Lovakov & Agadullina, 2017).

Procedure. Procedures mirrored Phase 2 in Study 3 except participants rated mental images using an expanded set of 12 subjective feature dimensions that included more infectious disease threat and non-infectious disease threat dimensions: germy, disfigured, old, heavy, foreign, fatigued, healthy, violent, angry, dominant, muscular, and masculine.

Results

Analysis plan. Our analysis plan mirrored our plan from Phase 2 of Study 3.

Do Infected persons appear to have stronger infection-related features than less infection-related features? Similar to Study 3, comparing mean feature ratings within the Infected representation, I do not find sufficient evidence that the Infected representation appeared to have stronger infection-related features than less infection-related features (bottom left plot of Figure 8; see supplement for pairwise tests).

Do people discriminate between Healthy and Infected mental images? Using canonical discriminant analyses, I found that the twelve feature dimensions combined well to maximally discriminate between the Healthy and Infected mental images (cross-validated classification accuracy: 85%, 95% CI [79%, 89%]), Canonical R_2 = .43, $F(12, 197) = 12.22$, $p < .001$ (see Figure 14A). Specifically, raters judged the Infected mental image to be more germy (*d* = 0.81), more disfigured (*d* = 0.47), older (*d* = 0.85), heavier (*d* = 1.01), more foreign (*d* = 0.38),

more fatigued (d = 1.06), less healthy (d = -0.98), more violent (d = 0.68), and angrier (d = 0.84) than the Healthy mental image (see Figure 14B). Raters did not significantly distinguish the Infected representation from the Healthy representation on dominance (d = -0.06), muscularity (d = -0.06), or masculinity (d = -0.21). When considering these feature ratings differences and their correlations together, I observed Mahalanobis's D = 1.74, 95% CI [1.30, 1.95], which corresponds to 23.85% overlap (assuming multivariate normality). Importantly, I observed some heterogeneity: The equivalent proportion of variables coefficient (EPV) suggests 45.48% (about 5) of the feature dimensions contribute equally to the multivariate effect, D. Specifically, it seems that perceptions of germiness, heaviness, age, fatigue, health, violence, and anger contributed most strongly to configural differences in Healthy and Infected mental images.

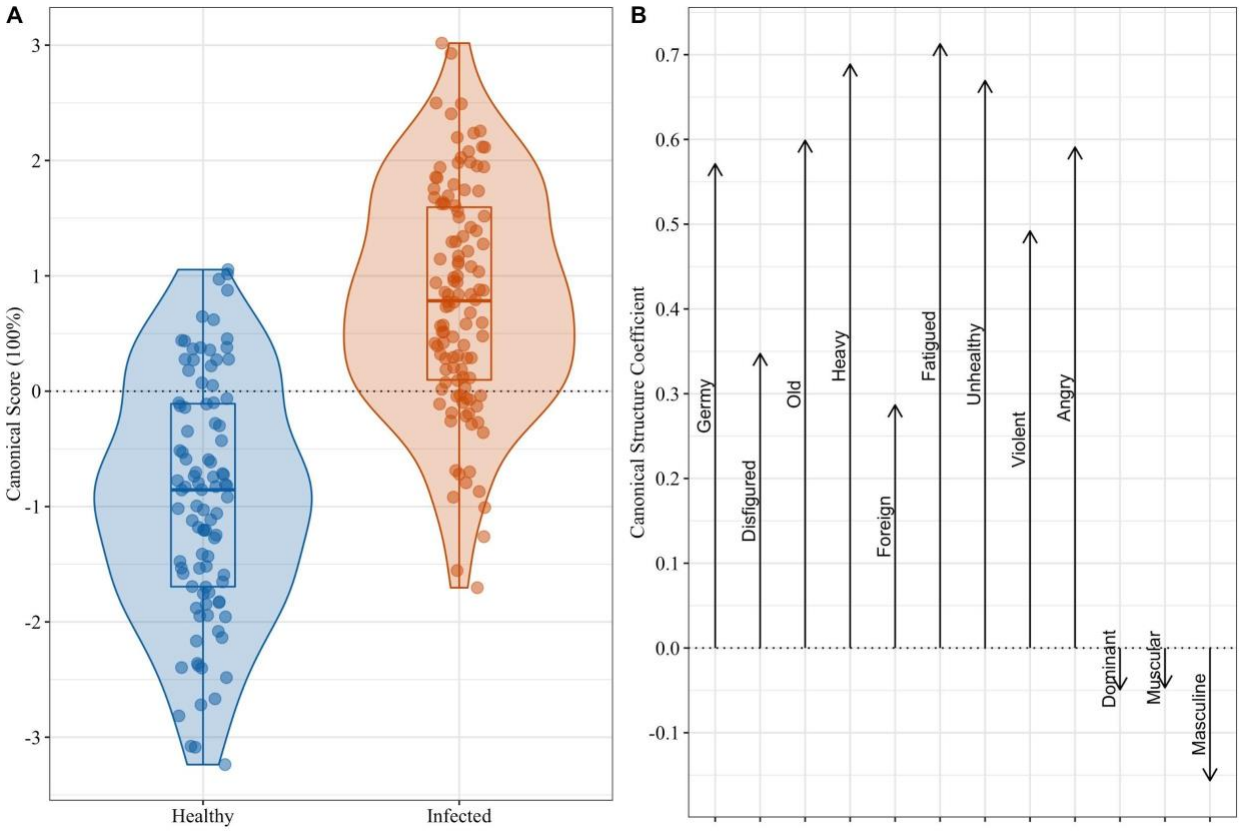

Figure 14. Reprinted from Michalak and Ackerman (2020). Together, the panels depict the maximal, configural difference between the Healthy and Infected classification images (Panel A) and the relative contribution of each rating to that difference (Panel B) (Study 4). The canonical scores represent participant ratings transformed to maximize the difference in the canonical variable between the Healthy and Infected conditions. Higher scores indicate a more Infected blend, and lower scores indicate a

more Healthy blend. Panel A depicts a combination of boxplots and violin plots that visualize representation canonical scores. Panel B depicts the direction and magnitude of partial (i.e., condition-adjusted) correlations between the individual feature dimension ratings and the canonical scores. Higher scores index the contribution of each feature dimension to the Healthy/Infected differences.

Do people want to avoid Infected persons more than Healthy persons? Raters reported wanting to avoid the person in the Infected mental image more, $M_{Difference}$ = 0.91, 95% CI [0.24, 1.57], t(201.82) = 2.68, p = .008, and they reported being willing to stand near the person in the Infected mental image less, $M_{Difference}$ = -1.00, 95% CI [-1.60, -0.40], t(199.39) = 3.27, p = .001, than the person in the Healthy representation. Thus, like Study 3, Study 4 raters wanted to avoid Infected people more than Healthy people.

Discussion

In Study 4, participants generated mental images of either Healthy or Infected people. An independent sample of participants rated these on a variety of features. Examining only the Infected representation, I did not find sufficient evidence that the Infected representation appeared to have stronger infection-related features than less infection-related features. However, when making multidimensional comparisons between face types, people did strongly distinguish between the Infected and Healthy representations; germiness, heaviness, age, fatigue, health, violence, and anger contributed most strongly to configural differences. Also, as in Study 3, the Infected representation appeared like someone people would want to avoid contact with. In sum, Infected representations had similar mean ratings across negative features, and Infected representations were judged to be more extreme than Healthy representations on both disease threat-specific features as well as on features not directly linked to infectious disease (violence and anger). Thus, as in Study 3, these ratings data are more consistent with the threat-combination hypothesis.

Notably, findings from the Infected category in Study 4 were consistent with those from the Germy category in Study 3 in that participants distinguished these composites from their respective comparison categories (non-Germy and Healthy) on the same feature dimensions of germiness, disfigurement, foreignness, heaviness, age, and violence. These data suggest that Infected representations largely overlap with Germy representations (and Healthy with non-Germy representations), and they are inconsistent with the interpretation that the Study 3 results were an artifact of comparing anti-classification images to classification images. This interpretation is further supported by objective correlations among the pixel luminance values in the images themselves: The Study 3 Germy composite strongly, positively correlated with the Study 1B Infected composite ($r = .60$) and strongly, negatively correlated with the Healthy composite ($r = -.58$), even though all 3 composites were generated in independent reverse correlation image classification tasks.

Study 5

So far, data-driven mental images from Studies 3 and 4 indicate that mental images of infected people include features previously linked to infectious disease threat as well as some features not directly associated with this threat. These data-driven representations appear to more strongly support the threat-combination hypothesis over the threat-specific hypothesis, unlike images generated through expectation-driven methods. Though data-driven representations may include a combination of threatening features, it is also possible that threat-specificity emerges more clearly when contrasting representations linked to different types of threats. For example, it may be that representations of infected persons appear more germy but less aggressive than representations of violent persons. Such representations may possess unique patterns of cues indicating that these people pose qualitatively different threats.

To examine this possibility, in Study 5 participants chose faces best matching either Germy or Violent categories. "Germy" represents a person who poses an infectious disease threat (as in earlier Studies), and "Violent" represents a person who poses a physical harm threat. I then compared classification images between these two categories, as in Study 4. All other aspects of the study design followed those used in prior studies.

Method

Phase 1: Estimating mental images of Violent and Germy persons

Participants. I planned to recruit 200 participants (n = 100 for each target category) using TurkPrime (Litman et al., 2017) (see preregistration at https://aspredicted.org/wa5mj.pdf). Between February 9th, 2017 and February 10th, 2017, I recruited participants until I had approximately equal numbers of participants between conditions and $N \geq 200$. Participants opened our survey 398 times and I paid $2.00 to all 244 (61.46% completion rate) participants who submitted their MTurk HIT assignments (see pertinent sample characteristics in Table 3). I randomly assigned these participants to a Violent (n = 102) or Germy (n = 98) condition. Statistical power. The condition with the smallest sample size (n = 98) afforded 80% power to detect Pearson's r = .27 (pwr.r.test function in the pwr package in R; Champely, 2018).

Procedure. For the reverse correlation task, I used the same base face and noise patterns generated for Study 3. Like our procedure in Study 4, participants read instructions which included a definition of their randomly-assigned target category: (1) Violent, "prone to commit acts of violence; uses physical force intended to hurt, damage, or kill someone or something," or (2) Germy, "full of germs; germ infested; appearing either sick or contaminated." Next, they completed a practice trial. Before moving onto the task, participants saw their target definition again above a reminder describing which keys correspond to selecting the image on the left or

right during each trial. After completing the reverse correlation image classification task, students completed three trait-level threat concern questionnaires, reported what they thought was the purpose of the study, and saw a short debriefing page.

Generating classification images. Using rcicr, I created classification images and anti-classification images for chooser participants' mental images of Violent and Germy (see Figure 15). In total, I created one Violent and one Germy classification image as well as one Violent anti-classification image and one Germy anti-classification image. As in Study 4, I only report analyses on classification images.

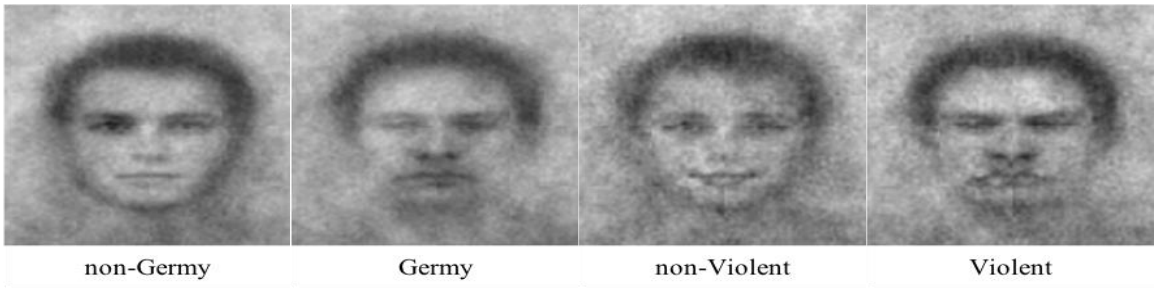

Figure 15. Reprinted from Michalak and Ackerman (2020). Images depict non-Germy (top left), Germy (top right), non-Violent (bottom left), and Violent (bottom right) classification images (Study 5).

Phase 2: Measuring feature dimensions of Violent and Germy mental images

Participants. I planned to recruit 500 MTurkers using TurkPrime (Litman et al., 2017) (see preregistration at https://aspredicted.org/q6wv2.pdf). Between June 13th, 2017 and June 20th, 2017, I sampled participants until I had recruited approximately equal numbers of MTurkers among four conditions and $N \geq 500$. Participants opened our survey 726 times and I paid $0.75 to all 542 (74.66% completion rate) participants who submitted their MTurk HIT assignments; 505/542 (93.17%) met our inclusion criteria (see Table 3 for pertinent sample characteristics). I randomly assigned these participants to rate a Violent anti-classification image (n = 124), a Violent classification image (n = 127), a Germy anti-classification image (n = 129),

or a Germy classification image (n = 126). As in Studies 3 and 4, participants also rated additional images assessing individual difference effects (see supplement).

Statistical power. Our final sample (N = 505) afforded us 80% power to detect Cohen's $d = 0.25$ for main effects and interactions and $d = 0.35$ for two-cell contrasts (power.t.test function in the stats package in R; (R Core Team, 2019), effect sizes near the estimated median in intergroup processes research (Lovakov & Agadullina, 2017).

Procedure. Procedures mirrored those from Phase 2 in Study 4.

Results

Analysis plan. Our analysis plan mirrored our plans from Phase 2 of Studies 3 and 4.

Do Germy persons appear to have stronger infection-related features than less infection-related features? As in earlier studies, comparing mean feature ratings within the Germy representation, I do not find sufficient evidence that the Germy representation appears to have stronger infection-related features than less infection-related features (bottom right plot of Figure 8; see supplement for pairwise tests).

Do people discriminate between Violent and Germy mental images? Using canonical discriminant analyses, I found that the twelve subjective feature dimensions combined well to maximally discriminate between the Violent and Germy mental images (cross-validated classification accuracy: 72%, 95% CI [65%, 77%]), Canonical $R^2 = .22$, $F(12, 240) = 5.60$, $p < .001$ (see Figure 16A). Specifically, raters judged the Germy mental image to be less old ($d = -0.28$), less foreign ($d = -0.72$), less violent ($d = -0.59$), less angry ($d = -0.68$), less dominant ($d = 0.60$), less muscular ($d = -0.50$), and less masculine ($d = -0.34$) than the Violent mental image (see Figure 16B). Raters did not judge the Germy representation as significantly more germy ($d = -0.09$), disfigured ($d = -0.12$), heavy ($d = -0.23$), fatigued ($d = 0.21$), or healthy ($d = -0.14$) than

the Violent representation. When considering these trait differences and their correlations together, I observed Mahalanobis's D = 1.05, 95% CI [0.77, 1.20], which corresponds to 42.70% overlap (assuming multivariate normality). Importantly, I observed some heterogeneity: The equivalent proportion of variables coefficient (EPV) suggests 36.61% (about 4) of the feature dimensions contribute equally to the multivariate effect, D. Specifically, it seems that perceptions of foreignness, violence, anger, and dominance contributed most strongly to configural differences in Violent and Germy mental images.

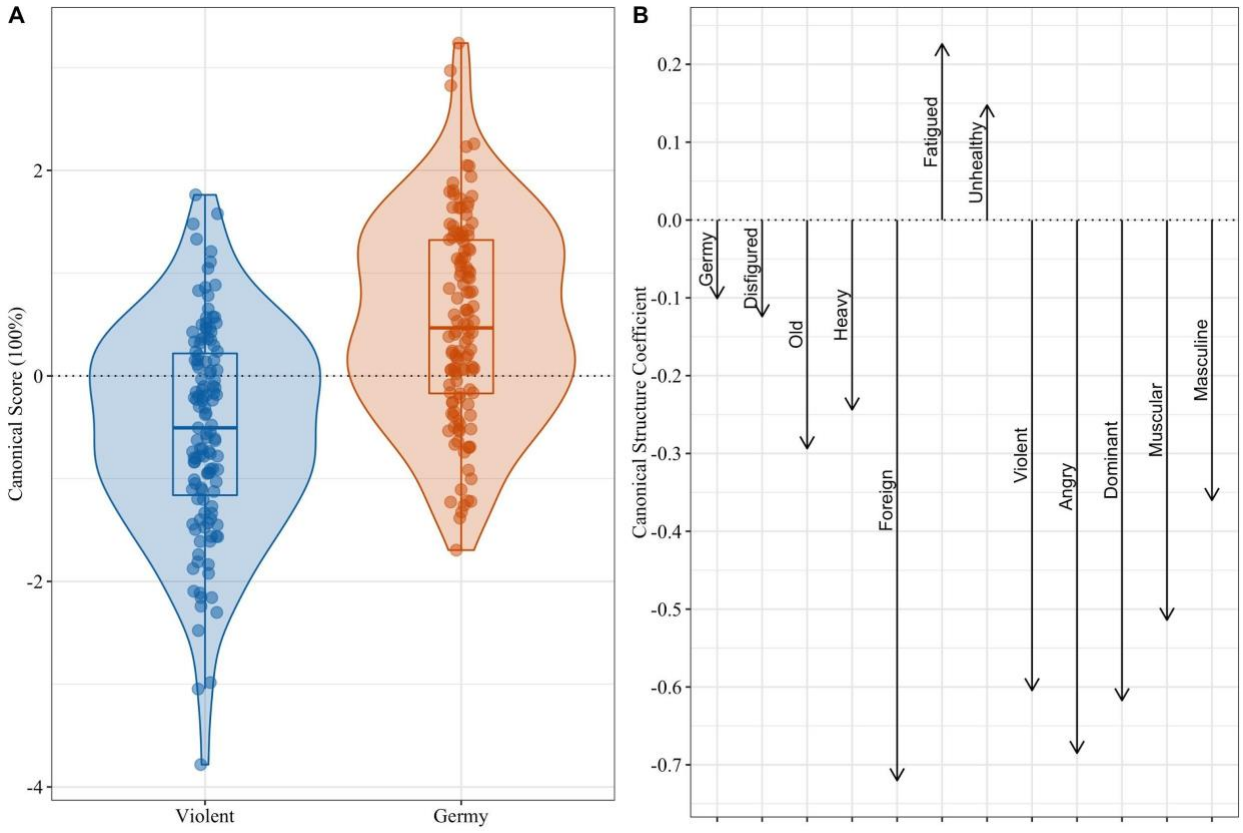

Figure 16. Reprinted from Michalak and Ackerman (2020). Together, the panels depict the maximal, configural difference between the Violent and Germy classification images (Panel A) and the relative contribution of each rating to that difference (Panel B) (Study 5). The canonical scores represent participant ratings transformed to maximize the difference in the canonical variable between the Violent and Germy conditions. Higher scores indicate a more Germy blend, and lower scores indicate a more Violent blend. Panel A depicts a combination of boxplots and violin plots that visualize representation canonical scores. Panel B depicts the direction and magnitude of partial (i.e., condition-adjusted) correlations between the individual feature dimension ratings and the canonical scores. Higher scores index the contribution of each feature dimension to the Violent/Germy differences.

Do people want to avoid Germy persons more than Violent persons? I found no sufficient evidence that raters wanted to avoid the person in the Germy mental image more, 95% CI [-1.10, 0.10], or stand near them less, 95% CI [-0.44, 0.67], than the person in the Violent mental image. Thus, the composites appear to be perceived as equally threatening, in general.

Discussion

As in Studies 3 and 4, I found no sufficient evidence that the Germy representation appeared to have stronger infection-related features than infection-unrelated features. However, unlike Studies 3 and 4 in which I compared mental images of people associated with infectious disease threat to representations of people who presumably pose no threat (i.e., non-Germy, Healthy), in Study 5, I compared a representation of infection threat to a representation of a violence threat. This comparison afforded us a complementary test of our hypotheses.

Among the subjective feature dimensions measured, perceptions of foreignness, violence, anger, and dominance contributed most strongly to configural differences between Violent and Germy mental images. These contrasts suggest Germy mental images—when compared to Violent mental images—appear to share many of the same disease threat-specific feature dimensions studied in the pathogen avoidance literature (e.g., germy, healthy, disfigured). This is indexed in part by lower classification accuracy by the canonical discriminant analysis (compare 78% and 85% accuracy when distinguishing infection threat from no-threat in Studies 3 and 4 to 72% when distinguishing infection threat from violence threat in Study 5). From the threat-specific perspective, this pattern is surprising. Though the Germy representations do possess

infectious disease features[4], Violent representations also possess such features. In contrast, Germy representations appear to pose less of a physical harm threat relative to Violent representations (e.g., less violent, angry, dominant). Thus, Germy representations are marked by infectious disease threat-specific features, but these features do not appear to be more pronounced than the same features in the Violent representations. Also, even though the Violent representation appeared to pose a stronger physical harm threat, I found no sufficient evidence that raters wanted to avoid such a person more than the person represented in the Germy image.

The partial overlap between features associated with both Germy and Violent representations is further supported by objective correlations between their pixel luminance values. The Violent image correlates less strongly with the Germy image from Study 5 ($r = .45$) and Study 3 ($r = .32$) than the Study 5 Germy image correlates with the Infected image from Study 4 ($r = .65$) and the Germy image from Study 3 ($r = .60$).

Taken together, these data suggest people represent both Germy and Violent persons with features from multiple specific threats, but Violent representations possess more prominent, physical harm threat-specific features. I interpret these patterns as being most consistent with the threat-combination hypothesis.

General Discussion

I started this research with a simple question: How do people mentally represent distinct interpersonal threats? I proposed two hypotheses. The threat-specificity hypothesis predicts that people mentally represent distinct interpersonal threats with features specific to and diagnostic of

[4] Whether we compare Germy representations to non-Germy representations (Study 3) or Infected to Healthy representations (Study 4), the infectious disease threat representation is consistently rated more germy, more disfigured, more fatigued, and less healthy.

those threats. Alternatively, the threat-combination hypothesis predicts that people mentally represent distinct threat categories with a combination of threat features common across multiple types of threats. In tests of these hypotheses using different methodological approaches, I found that expectation-driven approaches—those privileging participant beliefs, stereotypes, and intuitions—revealed strong evidence of threat-specificity. In contrast, data-driven approaches—those constraining top-down processes—revealed evidence more consistent with the threat-combination hypothesis. For example, participants who listed traits or drew images of their representations associated infection threat with infection-specific features and violence threat with aggression-specific features. But when participants produced mental images through the reverse correlation task, both threat representations included infection- and violence-relevant features (though the magnitude of these features differed across categories). These patterns suggest two key takeaways: (1) consistent with a functional perspective, threat-specificity emerges in representations most strongly when perceivers can control the content of their responses through editing, stereotype application, and so on, and (2) method matters.

Implications

The patterns uncovered here have implications for our understanding of pathogen avoidance psychology as well as threat management more generally. Research on how people process and react to others in the context of pathogen threat has commonly focused on a small set of cues that elicit negative interpersonal responses, such as unfamiliarity (e.g., foreignness, outgroupness; (Aarøe et al., 2017; Faulkner et al., 2004) and aspects of physical appearance (e.g., obesity, age, unattractiveness; Duncan & Schaller, 2009; Park et al., 2007; Tybur & Gangestad, 2011). The study of such cues is premised on functional perspectives that emphasize the specific costs and consequences of infection relative to other types of threat (Neuberg et al., 2011).

Additionally, many of these research findings stem from participant reactions to researcher-selected threat cues. Yet, our findings suggest that mental images of infected persons generated through data-driven methods include features that are not specific to pathogen threat (e.g., anger). Such results suggest that approaches drawing on functional (and other) perspectives could benefit from use of more diverse methods. The fact that representations were more distinct when responses were unconstrained indicates participants expect threat-specificity much like functional threat management researchers often do. An implication is that research that focuses too narrowly on threat-specific cues may overlook the complex ways in which people represent and understand these hazards.

This social cognitive and perceptual approach to infection threat also raises questions about the downstream consequences of holding mixed-threat mental representations. For instance, perceivers may draw negative inferences about others displaying cues associated with infection not only because perceivers anticipate the potential for infection but also because they infer the potential for other harms. Such inferences could influence the stereotypes perceivers apply (e.g., broadly negative), attributions they make (e.g., difficult, hostile), and behavioral responses (e.g., avoid) to these people. Outside the lab, these threat management processes could affect how sick people (and those merely resembling sick people) are treated.

Last, our findings speak to how people measure and conceptualize mental representations of other social categories. Researchers study social categories using a variety of methods, including the methods used here. However, they do not always explicitly consider whether different methods can privilege different psychological processes, thereby affecting their conclusions about how people represent social categories. Given I found participants emphasize different features in their mental representations depending on their task, it is possible that

mental representations of other social categories (including ones not associated with threat) might "look" different depending on the methods researchers use to estimate them. Perhaps the best solution to this challenge is the use of multiple methods whose strengths compensate for the limitations of other methods, providing a more holistic view on representations.

Limitations and future directions

The methods I employed here have many strengths such as the fact participants themselves constructed interpersonal threat representations rather than merely responding to representations provided by researchers. And the use of both expectation-driven and data-driven methods allows us to evaluate the influence of top-down processes on the generation of mental representations. However, these methods are limited in certain ways.

First, the reverse correlation image classification task entails trade-offs in process and final image generation. During this task, participants make all their selections first, and then researchers use software to average those selections. This average serves as a proxy mental image. Because participants do not see this mental image emerge as they make selections, they cannot adjust their representation as they see fit over the course of the task. Some may view this task feature as a limitation because mental representations emerge dynamically and draw from multiple sources of information in the mind, including salient stereotypes (Carlston & Smith, 1996; Freeman & Ambady, 2011; Sherman, 1996; Wyer, 2007). In our case, I found it useful to compare representations with and without this constraint feature.

Perhaps more importantly, reverse correlation classification images are created by averaging pixel patterns. Averaging is likely to obscure asymmetrical features that are hypothesized to reflect lower immunocompetence and, therefore, greater infection risk (Thornhill & Gangestad, 2006). Averaging pixel patterns may also blur random variations that represent

skin blotches, sores, or other skin anomalies associated with infectious disease. Does this issue invalidate the use of the reverse correlation task for estimating mental images of infected others? Not necessarily. Consider the responses given when participants used an expectation-driven procedure in Study 1. Participants listed features such as drowsy eyes, tired, weak, and sick, all of which can be perceived in principle (and qualitatively) in the composite images from Studies 3-4. High-level features like these emerge from specific patterns of eye, mouth, and eyebrow configurations, and the reverse correlation task is well suited to recovering such features. Thus, although this task cannot generate representations with certain features, our finding that the images generated from this task differed from images generated by methods like active drawing need not imply a methodological problem. This difference simply highlights that expectation-driven methods can limit our understanding of people's representations.

An additional concern about the reverse correlation task is that participants could simply select whichever images they perceived more negatively rather than selecting the images they perceived as more infected or violent per se. If so, the final classification images are averages of negative features rather than threat-specific features. I propose the data suggest otherwise. A negative feature choice rule should produce classification images with negative traits in roughly equal magnitude within and between threat representations. This was not the case. Within the germy/infected category, mental images included statistically distinguishable feature differences (not all negative traits were the same). The pattern of ratings and their correlations strongly distinguished the infected and violent representations, even though both categories appeared negative along many dimensions. In addition, certain feature ratings more strongly contributed to differences between categories (e.g., fatigue and healthy, Study 4), and the pixel luminance values of the germy and infected composites were more strongly correlated with each other than

with the violent images, across studies. Thus, rather than generate indiscriminately negative images, participants generated images with distinct patterns of negative features.

Although the expectation-driven methods used in Studies 1 and 2 may appear less limited than the reverse correlation task, they too have certain limitations. I detailed a primary one earlier—expectation-driven methods allow participants to edit their representations, which may look very different compared to their spontaneous representation. In addition, expectation-driven methods are susceptible to social desirability bias. For example, participants may associate a specific group with infection or germs but craft an image that does not reveal this association to researchers.

Future research might overcome the limitations of both the reverse correlation image classification task and the drawing method by using a task that provides on-the-fly representation updating by dynamically constructing a composite after each participant face choice and allowing participants to view this composite prior to the next face choice. In this way, a variety of hypothesized asymmetries (Thornhill & Gangestad, 2006) and skin anomalies could emerge in representations rather than blend together in a final average image. I are not aware of such a task, but I expect that this would be valuable to researchers interested in studying such cues.

Other limitations of our methods stem from stimuli-specific design decisions. Following early research on the reverse correlation image classification task (e.g., Dotsch et al., 2008, I used a grayscale base image that averages over a variety of White men making neutral expressions (Lundqvist et al., 1998). Thus, participants in our studies could not generate mental images of germy, infected, or violent people who were non-white, female, or in color. Each of these choices produces limitations that could be addressed in future research. One might expect that associations between certain demographic categories and threats produce somewhat different

mental representations than those found here. For instance, do some perceivers represent infected people with male features more than female features (as many infectious diseases are more prevalent in men; van Lunzen & Altfeld, 2014)? Are perceivers more likely to include other types of threat features in their representations of infection when faces depict groups stereotypically associated with certain threats (e.g., Black men are often stereotyped as aggressive in the United States; (Devine, 1989; Hugenberg & Bodenhausen, 2003)?

The use of color images in the reverse correlation task is a particularly interesting avenue for future work. Although greyscale base images are standard in the literature (Lundqvist et al., 1998), mental representations of infected people may commonly include color cues. For example, infected person representations may contain red pox or yellow-tinged skin or eyes (i.e., jaundice). Future research could allow reverse correlation stimuli to vary along relevant color dimensions or test questions about how people mentally represent color in infected others. Toward this goal, Gill et al., (2015) are developing a task that incorporates color using a "Bubble Warp" approach, using random, colored image fragments rather than greyscale faces overlaid with random noise (Gosselin & Schyns, 2001). Including color in the face could make a meaningful difference in healthy and infected appearance, and such color deviations (e.g., pale skin, red skin patches) could even help distinguish an infected representation from another threat category.

Conclusion

Over the course of human history, a variety of interpersonal threats like communicable diseases and interpersonal violence have posed strong selection pressures that favored the development of psychological systems that help people identify and ultimately reduce these threats. In this research, I examined whether the same cues that prior research has shown people

associate with infection and violence threats emerge in their mental representations of those threats. I found evidence that threat-specific cues do characterize these representations, but this was primarily true when the mental representation task allowed participants to deliberate on and edit their representations, privileging their expectations. In contrast, when a data-driven reverse correlation task constrained the influence of such expectations, mental images appeared distinguishable but with combinations of cues common across multiple distinct threat categories. Thus, the methods we employ to measure types of mental representations may shape the conclusions we draw about those representations. A multi-method approach may be our best option for fully capturing how people represent threat in their mind's eye.

Chapter IV. Discussion

Humans have been battling harmful pathogens for a long time, long enough to have developed sophisticated physical, psychological, and behavioral defenses against them. The physical immune system possesses remarkable machinery for identifying invading pathogens and eliminating them from the body, but its metabolic costs are too high to rely on it as the only defense mechanism. Humans have likely evolved another line of defense to reduce such costs: a behavioral immune system capable of identifying and avoiding pathogen infections in the first place. In chapter one, I laid out an evolutionary argument for the behavioral immune system as well as empirical evidence for one of its main hypotheses: the over-perception hypothesis. That is, given uncertainty inherent in identifying infection risks from imperfect cues and asymmetric costs in identification errors (i.e., false negatives are probably more costly), the behavioral immune system evolved a functional bias toward false positive errors: perceiving infection in objectively non-infectious cues. In chapters two and three, I investigated potential social perception mechanisms of the behavioral immune system that expand on this over-perception hypothesis.

In chapter two, I introduced the idea of proper and actual domains of the behavioral immune system. To review, researchers argue that functional mental modules like the behavioral immune system possess specialized functions with restricted inputs. Theoretically, the behavioral immune system evolved to process cues that correlate with infection—its proper domain of inputs (e.g., rashes, pustules, swelling). Given the design of its perceptual mechanisms, the behavioral immune system is capable of processing cues that share perceptual features with true

infection indicators (e.g., port-wine stain birthmarks, acne, obesity). The set comprising these cues and the proper domain cues constitute the actual domain.

In chapter two, I tested the hypothesis that people more strongly associate infection concepts with facial disfigurement than with obesity because facial disfigurement more closely resembles proper domain cues of the behavioral immune system than obesity does. I found support for this hypothesis across three studies that varied the nature of the comparison (between-subjects vs. within-subjects) and whether participants knew the facial disfigurement was benign. Importantly,in a fourth study, I did not find support for an alternative hypothesis that people associate any negative concepts more strongly with facial disfigurement than with obesity (e.g., laziness). All together, these results are consistent with the hypothesis that facial disfigurement more closely resembles a proper domain cue of the behavioral immune system than obesity does.

In chapter three, I used a variety of methods to investigate how people mentally represent infected others. I proposed two hypotheses for how these mental representations may appear. First, I proposed the threat-specificity hypothesis: people represent infected others with features that diagnose infection. Second, I proposed the threat-combination hypothesis: people mentally represent infected others with a variety of threatening features, including features that correlate with infection. Support for these hypotheses depended on the method I used to measure mental representations. The trait-listing and drawing methods yielded representations that appeared more consistent with the threat-specificity hypotheses, whereas the reverse correlation classification images yielded representations that appeared more consistent with the threat-combination hypothesis. All together, the results are consistent with the notion that methods that incorporate researcher or participant expectations (e.g., infected people should only look

infected) are more likely to yield results in line with those expectations. It turned out that participants in my studies expected infected people to have infection-specific features. But when they could not easily edit their representation in line with their expectations during the reverse correlation classification task, their aggregate choices appeared to have threatening features less diagnostic of infection (e.g., representations appeared angry and dominant).

Limitations and Future Directions

The research I presented in my dissertation yields insights into psychological mechanisms of the behavioral immune system, but the research designs I used limited the conclusions I could draw from the results. In chapter two, I claimed that facial rashes are a proper domain cue of the behavioral immune system and obesity is not, so people should more strongly associate infection with facial rashes and cues that resemble facial rashes than with obesity. Proper domain cues for the behavioral immune system should have a relatively long history of diagnosing infection in people. This is probably the case for facial rashes and less the case for obesity or even facial swelling (Wolfe et al., 2007) (though this premise needs more empirical support). Thus, the observed stronger association between infection concepts and facial disfigurement than with obesity is consistent with the hypothesis that facial disfigurement resembles a proper domain cue and obesity is not, or perhaps both cues resemble proper domain cyes but facial disfigurement resembles are more diagnostic, less noisy proper domain cue. However, the results are also consistent with the hypothesis that people learn associations between infection concepts and a variety of cues over their lifetime (i.e., cultural learning). A stronger test of this hypothesis would specify when during the life course humans might develop pathogen detection mechanisms (e.g., when they're able to avoid potentially infected people themselves). People should more strongly associate infection with proper domain cues during

and after this developmental stage. Similarly specified evolutionary hypotheses have been proposed for different psychological systems, such as the development of food neophobia, attachment and reproductive strategies, emotion, and danger learning ((Cashdan, 1998; Clark Barrett et al., 2016; Del Giudice, 2009; Frankenhuis, 2019; Wright, 1991). In general, aside from measuring disgust sensitivity in children, researchers have not tested behavioral immune system hypotheses in infants or children. It remains an open question how the behavioral immune system develops and operates over the life course.

In chapter three, I claimed to have measured people's mental representations of infected and violent others. In my general discussion in that chapter, I described important sampling characteristics that limited my findings as well as many findings in the behavioral immune system literature (Henrich et al., 2010). Despite that limitation, chapter three serves as a rough template for how behavioral immune system researchers as well as threat management researchers more broadly might develop mental representations as constructs. Researchers in these subfields develop and test a variety hypotheses about traits and motivations associated with categories of people: infected people, violent people, romantic partners, leaders, coalition members, friends, cheaters, parents, children, and so on (Boyer et al., 2015; Delton et al., 2012; Karremans et al., 2011; Krems & Conroy-Beam, 2020; Kurzban et al., 2001; Li et al., 2019; Neuberg et al., 2011; Van Vugt et al., 2008). In many cases, researchers are not necessarily interested in how social categories are represented but in how people think or feel about them or behave toward them. However, obviously, social psychological processes and behaviors depend on how social categories are perceived and represented, so it is worthwhile to study the representations themselves. How? As my research in chapter 3 demonstrates, what one concludes about how people perceive a social category depends on how you measure it. There's no single

method in the social sciences that captures how people represent or perceive a social category. People may self-report their social category representations differently than they draw them and differently than patterns extracted from categorization tasks they complete (e.g., classification images from reverse correlation tasks). Researchers would benefit from developing social category representations with the same theoretical and exploratory rigor that they use to develop other constructs (Campbell, 1960; Cronbach & Meehl, 1955).

Conclusion

In my dissertation, I investigated two broad psychological phenomena of the behavioral immune system: the proper domain of the behavioral immune system and mental representations of infected others. First, I observed initial evidence that people associate infection concepts more strongly with some anomalous yet benign facial features than with others, theoretically because the former more closely resembles cues the behavioral immune system was designed to process—its proper domain. Second, I used a variety of methods to measure how people mentally represented infected others and found that people listed traits and drew infected people with predominantly infection-related features (e.g., unhealthy, disfigured) but their reverse correlation classification images appeared with both infection-related and infection unrelated features (e.g., angry, dominant). These results yield novel insights into the mental mechanisms of the behavioral immune system.

References

Aarøe, L., Petersen, M. B., & Arceneaux, K. (2017). The Behavioral Immune System Shapes Political Intuitions: Why and How Individual Differences in Disgust Sensitivity Underlie Opposition to Immigration. *American Political Science Review*, *111*(2), 277–294. https://doi.org/10.1017/S0003055416000770

Ackerman, J. M., Becker, D. V., Mortensen, C. R., Sasaki, T., Neuberg, S. L., & Kenrick, D. T. (2009). A pox on the mind: Disjunction of attention and memory in the processing of physical disfigurement. *Journal of Experimental Social Psychology*, *45*(3), 478–485.

Ackerman, J. M., Hill, S. E., & Murray, D. R. (2018). The behavioral immune system: Current concerns and future directions. *Social and Personality Psychology Compass*, *12*(2), e12371. https://doi.org/10.1111/spc3.12371

Andersen, S. M., & Klatzky, R. L. (1987). Traits and social stereotypes: Levels of categorization in person perception. *Journal of Personality and Social Psychology*, *53*(2), 235–246.

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control*. Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, H. C. (2012). A hierarchical model of the evolution of human brain specializations. *Proceedings of the National Academy of Sciences*, *109*(Supplement 1), 10733–10740. https://doi.org/10.1073/pnas.1201898109

Barrett, H. C., & Kurzban, R. (2006). Modularity in Cognition: Framing the Debate. *Psychological Review*, *113*(3), 628–647. https://doi.org/10.1037/0033-295X.113.3.628

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boyer, P., Firat, R., & van Leeuwen, F. (2015). Safety, Threat, and Stress in Intergroup Relations: A Coalitional Index Model. *Perspectives on Psychological Science*, *10*(4), 434–450. https://doi.org/10.1177/1745691615583133

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science*, *28*(1), 92–103.

Brown-Iannuzzi, J. L., McKee, S., & Gervais, W. M. (2018). Atheist horns and religious halos:

Mental representations of atheists and theists. *Journal of Experimental Psychology: General*, *147*(2), 292–297. https://doi.org/10.1037/xge0000376

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, *15*(8), 546–553. https://doi.org/10.1037/h0048255

Carlston, D. E., & Smith, E. R. (1996). Principles of mental representation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 184–210). Guilford Press.

Cashdan, E. (1998). Adaptiveness of food learning and food aversions in children. *Social Science Information*, *37*(4), 613–632.

Champely, S. (2018). *pwr: Basic Functions for Power Analysis* (1.2-2) [R]. https://CRAN.R-project.org/package=pwr

Clark Barrett, H., Peterson, C. D., & Frankenhuis, W. E. (2016). Mapping the cultural learnability landscape of danger. *Child Development*, *87*(3), 770–781.

Coetzee, V., Perrett, D. I., & Stephen, I. D. (2009). Facial adiposity: A cue to health? *Perception*, *38*(11), 1700–1711.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. *Mapping the Mind: Domain Specificity in Cognition and Culture*, *853116*.

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology*, *66*(5), 882.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281.

Curtis, V., Aunger, R., & Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_4), S131–S133.

Del Giudice, M. (2009). Sex, attachment, and the development of reproductive strategies. *Behavioral and Brain Sciences*, *32*(01), 1–21. https://doi.org/10.1017/S0140525X09000016

Del Giudice, M. (2017). Heterogeneity coefficients for Mahalanobis' D as a multivariate effect size. *Multivariate Behavioral Research*, *52*(2), 216–221.

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PloS One*, *7*(1), e29265. https://doi.org/10.1371/journal.pone.0029265

Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: Dissecting the architecture of a moral concept. *Journal of Personality and Social Psychology*, *102*(6), 1252–1270. https://doi.org/10.1037/a0027026

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18.

Dobson, A. P., & Carper, E. R. (1996). Infectious diseases and human population history.

Bioscience, *46*(2), 115–126. https://doi.org/10.2307/1312814

Dotsch, R. (2016). *rcicr: Reverse correlation image classification toolbox* (0.3.4.1) [R]. https://cran.r-project.org/web/packages/rcicr/index.html

Dotsch, R., & Todorov, A. (2012a). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., & Todorov, A. (2012b). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Drage, L. A. (1999). Life-threatening rashes: Dermatologic signs of four infectious diseases. *Mayo Clinic Proceedings*, *74*(1), 68–72.

Duncan, L. A., & Schaller, M. (2009). Prejudicial attitudes toward older adults may be exaggerated when people feel vulnerable to infectious disease: Evidence and implications. *Analyses of Social Issues and Public Policy*, *9*(1), 97–115. https://doi.org/10.1111/j.1530-2415.2009.01188.x

Duncan, L. A., Schaller, M., & Park, J. H. (2009). Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. *Personality and*

Individual Differences, *47*(6), 541–546. https://doi.org/10.1016/j.paid.2009.05.001

Falagas, M. E., & Kompoti, M. (2006). Obesity and infection. *The Lancet Infectious Diseases*, *6*(7), 438–446.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, *95*(3), 307–317.

Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, *7*(4), 333–353. https://doi.org/10.1177/1368430204046142

Fessler, D. M., Clark, J. A., & Clint, E. K. (2015). Evolutionary psychology and evolutionary anthropology. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (2nd ed., Vol. 2, pp. 1029–1046). Wiley Online Library.

Fessler, D. M., Holbrook, C., & Snyder, J. K. (2012). Weapons make the man (larger): Formidability is represented as size and strength in humans. *PloS One*, *7*(4).

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, *32*(1), 1–24.

Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with

predictor effect plots and partial residuals. *Journal of Statistical Software, 87*(9), 1–27.

Frankenhuis, W. E. (2019). Modeling the evolution and development of emotions. *Developmental Psychology, 55*(9), 2002.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247–279. https://doi.org/10.1037/a0022327

Friendly, M., & Fox, J. (2020). *candisc: Visualizing Generalized Canonical Discriminant and Canonical Correlation Analysis* (0.8-3) [Computer software]. https://CRAN.R-project.org/package=candisc

Gill, D., DeBruine, L., Jones, B., & Schyns, P. (2015). Bubble-Warp: A new approach to the depiction of high-level mental representation. *Journal of Vision, 15*(12), 420–420.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A new technique to reveal the use of information in recognition tasks. *Journal of Vision, 1*(3), 333–333. https://doi.org/10.1167/1.3.333

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81–91.

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10*(1), 47–66.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The Evolution of Cognitive Bias. In *The*

Handbook of Evolutionary Psychology (pp. 1–20). John Wiley and Sons, Inc. https://doi.org/10.1002/9781119125563.evpsych241

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Holbrook, C., & Fessler, D. M. T. (2015). The Same, Only Different: Threat Management Systems as Homologues in the Tree of Life. In P. J. Carroll, R. M. Arkin, & A. L. Wichman (Eds.), *Handbook of Personal Security* (pp. 95–109). Psychology Press. https://doi.org/10.4324/9781315713595-12

Holbrook, C., Galperin, A., Fessler, D. M. T., Johnson, K. L., Bryant, G. A., & Haselton, M. G. (2014). If looks could kill: Anger attributions are intensified by affordances for doing harm. *Emotion*, *14*(3), 455–461. https://doi.org/10.1037/a0035826

Huang, J. Y., Sedlovskaya, A., Ackerman, J. M., & Bargh, J. A. (2011). Immunizing against prejudice: Effects of disease protection on attitudes toward out-groups. *Psychological Science*, *22*(12), 1550–1556.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*(6), 640–643.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*,

68(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Karremans, J. C., Dotsch, R., & Corneille, O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition, 121*(3), 422–426.

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

Krems, J. A., & Conroy-Beam, D. (2020). First tests of Euclidean preference integration in friendship: Euclidean friend value and power of choice on the friend market. *Evolution and Human Behavior*.

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 127*(2), 187–208. https://doi.org/10.1037/0033-2909.127.2.187

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Li, Y. J., Haws, K. L., & Griskevicius, V. (2019). Parenting motivation and consumer decision-making. *Journal of Consumer Research*, *45*(5), 1117–1137.

Lieberman, D. L., Tybur, J. M., & Latner, J. D. (2012). Disgust sensitivity, obesity stigma, and gender: Contamination psychology predicts weight bias for women, not men. *Obesity*, *20*(9), 1803–1814.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. Com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lovakov, A., & Agadullina, E. (2017). Empirically Derived Guidelines for Interpreting Effect Size in Social Psychology. *PsyArXiv*. https://doi.org/10.31234/osf.io/2epc4

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces*. Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135.

Mancuso, P. (2013). Obesity and respiratory infections: Does excess adiposity weigh down host defense? *Pulmonary Pharmacology & Therapeutics*, *26*(4), 412–419.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the

information employed for face classifications. *Cognitive Science*, *28*(2), 209–226.

Martin, D. (2016). *IAT: Cleaning and Visualizing Implicit Association Test (IAT) Data* (0.3) [Computer software]. https://CRAN.R-project.org/package=IAT

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics*, *5*(3), 269–287.

McKay, R., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, *31*(5), 309–319.

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, *14*(11), 1256–1265.

Michalak, N. M., & Ackerman, J. (2017). *A multi-method approach to measuring mental representations of threatening others*. https://doi.org/10.17605/OSF.IO/84VDP

Michalak, N. M., & Ackerman, J. M. (2020). A multimethod approach to measuring mental representations of threatening others. *Journal of Experimental Psychology: General*. https://doi.org/xge-xge0000781

Miller, S. L., & Maner, J. K. (2012). Overperceiving Disease Cues: The Basic Cognition of the Behavioral Immune System. *Journal of Personality and Social Psychology*, *102*(6), 1198–1213.

Navarrete, C. D., & Fessler, D. M. (2006). Disease avoidance and ethnocentrism: The effects of disease vulnerability and disgust sensitivity on intergroup attitudes. *Evolution and*

Human Behavior, *27*(4), 270–282.

Nesse, R. M. (2005). Natural selection and the regulation of defenses: A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior*, *26*(1), 88–105. https://doi.org/10.1016/j.evolhumbehav.2004.08.002

Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience & Biobehavioral Reviews*, *35*(4), 1042–1051. https://doi.org/10.1016/j.neubiorev.2010.08.011

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin*, *135*(2), 303.

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior*, *27*(2), 65–87.

Park, J. H., Schaller, M., & Crandall, C. S. (2007). Pathogen-avoidance mechanisms and the stigmatization of obese people. *Evolution and Human Behavior*, *28*(6), 410–414. https://doi.org/10.1016/j.evolhumbehav.2007.05.008

Pontzer, H., Raichlen, D. A., Wood, B. M., Mabulla, A. Z., Racette, S. B., & Marlowe, F. W. (2012). Hunter-gatherer energetics and human obesity. *PloS One*, *7*(7), e40503.

R Core Team. (2019). *R: A language and environment for statistical computing*. https://www.R-project.org/

Ratner, K. G., Dotsch, R., Wigboldus, D. H. J., van Knippenberg, A., & Amodio, D. M. (2014).

Visualizing minimal ingroup and outgroup faces: Implications for impressions, attitudes, and behavior. *Journal of Personality and Social Psychology*, *106*(6), 897–911. https://doi.org/10.1037/a0036498

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

Ryan, S., Oaten, M., Stevenson, R. J., & Case, T. I. (2012). Facial disfigurement is treated like an infectious disease. *Evolution and Human Behavior*, *33*(6), 639–646. https://doi.org/10.1016/j.evolhumbehav.2012.04.001

Schaller, M. (2016). The behavioral immune system. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (2nd ed., Vol. 1, pp. 206–224). Wiley. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-111875588X.html

Schaller, M., & Neuberg, S. L. (2012). Danger, disease, and the nature of prejudice (s). In *Advances in Experimental Social Psychology* (Vol. 46, pp. 1–54). Elsevier.

Schaller, M., Park, J., & Faulkner, J. (2003). Prehistoric dangers and contemporary prejudices. *European Review of Social Psychology*, *14*(1), 105–137.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology*, *70*(6), 1126–1141.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and

predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. *Mapping the Mind: Domain Specificity in Cognition and Culture*, 39–67.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, *8*(1), 40–46.

Stangor, C., & Lange, J. E. (1994). Mental representations of social groups: Advances in understanding stereotypes and stereotyping. In *Advances in experimental social psychology* (Vol. 26, pp. 357–416). Elsevier.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207–218.

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior*, *27*(2), 131–144. https://doi.org/10.1016/j.evolhumbehav.2005.06.001

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven Methods for Modeling Social Perception: Modeling Social Perception. *Social and Personality Psychology Compass*, *5*(10), 775–791. https://doi.org/10.1111/j.1751-9004.2011.00389.x

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, *19*.

Tybur, J. M., Frankenhuis, W. E., & Pollet, T. V. (2014). Behavioral immune system methods: Surveying the present to shape the future. *Evolutionary Behavioral Sciences, 8*(4), 274.

Tybur, J. M., & Gangestad, S. W. (2011). Mate preferences and infectious disease: Theoretical considerations and evidence in humans. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1583), 3375–3388. https://doi.org/10.1098/rstb.2011.0136

Tybur, J. M., & Lieberman, D. (2016). Human pathogen avoidance adaptations. *Current Opinion in Psychology*, *7*, 6–11. https://doi.org/10.1016/j.copsyc.2015.06.005

Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: Individual differences in three functional domains of disgust. *Journal of Personality and Social Psychology*, *97*(1), 103–122. https://doi.org/10.1037/a0015474

Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review*, *120*(1), 65–84.

van Leeuwen, F., Hunt, D. F., & Park, J. H. (2015). Is Obesity Stigma Based on Perceptions of Appearance or Character? Theory, Evidence, and Directions for Further Study. *Evolutionary Psychology*, *13*(3), 1474704915600565. https://doi.org/10.1177/1474704915600565

van Leeuwen, F., & Petersen, M. B. (2018). The behavioral immune system is designed to avoid infected individuals, not outgroups. *Evolution and Human Behavior*, *39*(2), 226–234. https://doi.org/10.1016/j.evolhumbehav.2017.12.003

van Lunzen, J., & Altfeld, M. (2014). Sex Differences in Infectious Diseases–Common but Neglected. *The Journal of Infectious Diseases*, *209*(suppl_3), S79–S80. https://doi.org/10.1093/infdis/jiu159

Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist*, *63*(3), 182–196. https://doi.org/10.1037/0003-066X.63.3.182

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer.

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences*, *55*(8), 994–999.

Westfall, J. (2016). PANGEA: Power analysis for general ANOVA designs. *Unpublished Manuscript. Available at Http://Jakewestfall.Org/Publications/Pangea.Pdf*.

Whigham, L. D., Israel, B. A., & Atkinson, R. L. (2006). Adipogenic potential of multiple human adenoviruses in vivo and in vitro in animals. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *290*(1), R190–R194.

Wolfe, N. D., Dunavan, C. P., & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, *447*(7142), 279–283.

Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods*, *49*(4), 1193–1209. https://doi.org/10.3758/s13428-016-0779-0

World Health Organization. (2018). *Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016*. World Health Organization. https://www.who.int/healthinfo/global_burden_disease/GHE2016_DALYs-2016-country.xls?ua=1

Wright, P. (1991). Development of food choice during infancy. *Proceedings of the Nutrition Society*, *50*(1), 107–113.

Wyer, R. S. (2007). Principles of mental representation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (Vol. 2, pp. 285–307). Guilford Press.

Zebrowitz, L. A., & Franklin Jr, R. G. (2014). The attractiveness halo effect and the babyface stereotype in older and younger adults: Similarities, own-age accentuation, and older adult positivity effects. *Experimental Aging Research*, *40*(3), 375–393.