A Multimethod Approach to Measuring Mental Representations of Threatening Others

Nicholas M. Michalak and Joshua M. Ackerman

A Multimethod Approach to Measuring Mental Representations of Threatening Others

Nicholas M. Michalak and Joshua M. Ackerman
University of Michigan

How do people mentally represent distinct interpersonal threats? Across human history, interpersonal threats such as infectious disease and violence have posed powerful selection pressures. Such pressures selected for psychological systems that help identify and reduce threats posed by other people. In the case of infectious disease, psychology researchers have found that such systems respond to a variety of infection cues (e.g., rashes, swelling) as well as cues that merely resemble infection cues (e.g., birthmarks, obesity). Are such cues part of people's mental representations, and if so, are those cues unique to infection representations or are they included in representations of other threats? Using a multimethod approach, we find that when participants listed traits or drew mental representations of threat, they perceived infected and violent others to differ along threat-specific features. However, when using a data-driven, reverse correlation method that restricted participants from deliberating on and editing their representations, participants generated mental images that were similar on many of the features that both researchers and laypeople expect to distinguish infection and violence threats. These findings suggest that our understanding of threat processing may suffer from a potential disconnect between the threat cues derived from the expectations of researchers and those revealed when expectations are constrained.

Keywords: behavioral immune system, evolutionary psychology, reverse correlation, threat management, violence

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.

We 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? We examine this question in the domain of pathogen threat psychology, where perceivers represent the faces of infected others.

We 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 we detail the strengths and limitations of expectation-driven versus data-driven methods of threat assessment. To preview our 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 we refer to as the functional threat management perspective, natural selection has favored mental systems that enable people to perceive, feel, think,

Nicholas M. Michalak and Joshua M. Ackerman, Department of Psychology, University of Michigan.
Supplemental Repository: https://osf.io/84vdp/.
Correspondence concerning this article should be addressed to Nicholas M. Michalak, Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109. E-mail: nickmm@umich.edu

and behave in ways to reduce threats in particular rather than threats in general (Barrett, 2012; Cosmides & Tooby, 1994; Tooby & Cosmides, 1992; Holbrook & Fessler, 2015; Neuberg, Kenrick, & Schaller, 2011). 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 with the same objects that had been touched by visibly healthy targets (Ryan, Oaten, Stevenson, & Case, 2012). People also expressed more disgust by and less willingness to touch those objects when the target people bore noninfectious 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, Hill, & Murray, 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, Holbrook, & Snyder, 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, Stevenson, & Case, 2009; Schaller, Park, & Faulkner, 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 & Schaller, 2009; Park, Schaller, & Crandall, 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. More than a decade of research found support for the hypothesis that people use perceptions of warmth and competence to understand social groups (Fiske, Cuddy, Glick, & Xu, 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, Imhoff, Dotsch, Unkelbach, & Alves, 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, 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, we 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, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997; Mechelli, Price, Friston, & Ishai, 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.

The Present Research

To address these limitations, we used multiple approaches to estimate mental representations of two threat categories: infected persons and violent persons. We evaluate results from each approach according to two hypotheses. Our first hypothesis—what we 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 we 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, we 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 & Lange, 1994; Stangor, Lynch, Duan, & Glas, 1992). 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, we 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, Todorov, & Dotsch, 2017; Dotsch & Todorov, 2012; Mangini & Biederman, 2004; Todorov, Dotsch, Wigboldus, & Said, 2011). For example, in the two-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 of unique categories, including racial and minimal out-groups (Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014), welfare recipients (Brown-Iannuzzi, Dotsch, Cooley, & Payne, 2017), and atheists (Brown-Iannuzzi, McKee, & Gervais, 2018). Unlike more common approaches, the reverse correlation method has received comparatively less attention in the threat processing literature. We 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 two-image forced choice image classification task, and in Phase 2, independent participants rated features of those images generated in Phase 1.

Overall, we report five studies using a multimethod approach. We 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, we preregistered research questions, predictions, sampling plans, exclusion criteria, and analyses. We 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, we 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/), we also include a spreadsheet detailing original and revised plans (i.e., those reported in this article). We also make available preregistrations, materials, and deidentified data for all studies, including data sets not reported here. Finally, we 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. 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

Participants. We recruited undergraduate psychology student participants between March 25th, 2019 and September 25th, 2019 (see Table 1 for pertinent sample characteristics).

Statistical power. We 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 nonintuitive arcsine transformation to obtain Cohen's h, which we report and explain next. Participants used a total of 3,335 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. e.g., 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. We 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 (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). We 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, we 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 1). 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 gray 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.

The most frequently used words in Infected trait responses were either not used in the Violent trait responses (pale and weak), used

Table 1
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 supplemental repository at https://osf.io/84vdp/ for additional information.

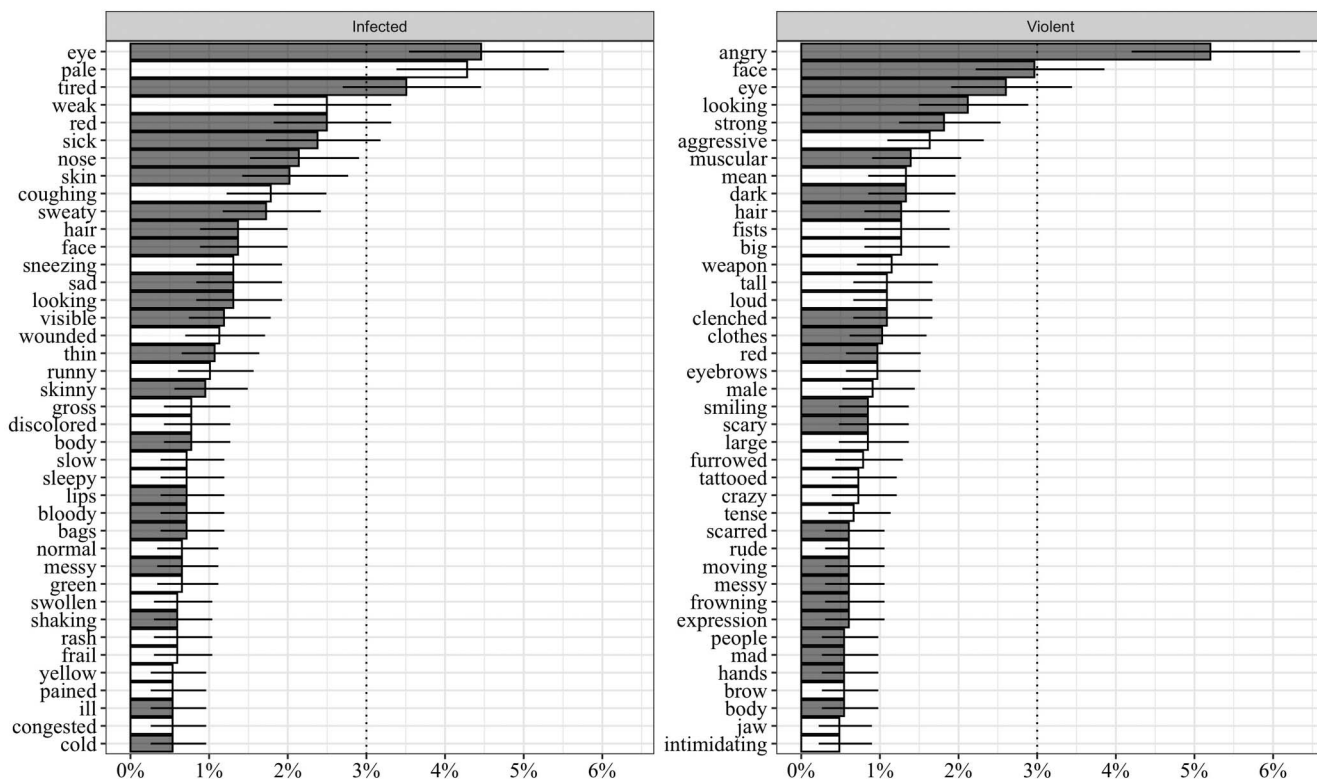## Word Frequencies (%) by Target Threat

Figure 1. 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 gray) 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. We added a dotted line at 3% for reference comparing across panels.

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 2). 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.

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. We asked participants to draw Germy (not Infected) and Violent persons in Phase 1, and then, in Phase 2, we 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: We

Table 2
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**

Note. Words are ordered by frequency within each threat category (1653 and 1682 words were used within violent and infected categories). The word *eye* topped both lists, so we table 19 instead of 20 most frequent words. Italics indicate Infected person words. *OR* represents the odds ratio; lower and upper together represent the 95% confidence interval limits for the *OR*; *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.

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. We recruited undergraduate psychology student participants between March 29, 2017 and April 7, 2017 (see Table 1 for pertinent sample characteristics).

Statistical power. We based power calculations 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. We initially recruited participants for a study focusing on separate research questions about infectious disease (see the supplemental 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 we 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."

We chose the *Germy* label after consulting with our undergraduate research assistants who believed that *Germy* would be most interpretable to undergraduate participants. See Figure 2 for example drawings (we make all our participants' drawings available in the supplemental repository). Among the 147 participants, 139 (94.56%) drew both a Germy and a Violent face. Our final stimuli sample comprised 139 pairs of drawings.[1]

Phase 2: Measuring subjective features of drawn faces.

Participants. We planned to recruit 138 MTurkers (assuming 10% would fail to meet inclusion criteria) using TurkPrime (Litman, Robinson, & Abberbock, 2017), leaving us with approximately 125 participants (see preregistration at https://osf.io/utyp6/). We sampled MTurkers until we had recruited approximately equal numbers of participants among 56 conditions and $N \geq 125$. One hundred forty-six participants opened our survey, and we paid $0.50 to all 143 (97.95% completion rate) participants who submitted their MTurk HIT assignments (see Table 1 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 versus Violent condition effect in the R(NCC) design described in (Judd, Westfall, & Kenny, 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, we entered into the Shiny Web application (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, we 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 rating portion of the study. Participants read definitions for 12 features 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

---

[1] By accident, 11 raters evaluated one 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.
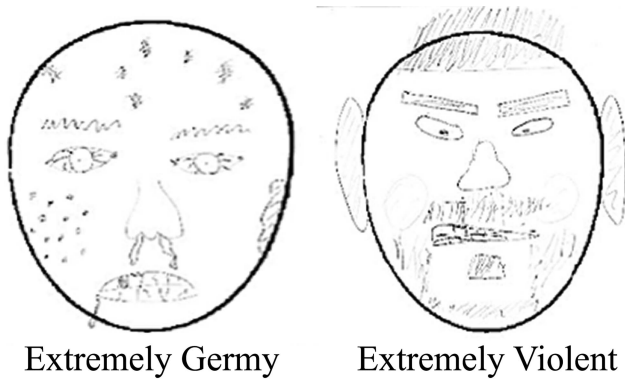
Figure 2. Images depict examples of a Germy drawing rated extremely Germy and a Violent drawing rated extremely violent (Study 2).

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, we followed analyses for the R(NCC) design recommended by (Judd et al., 2017). For each rating, we fit a linear mixed effects model using the lmer function from the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2014); Satterthwaite degrees of freedom and *p* values were calculated using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017). Specifically, we regressed feature rating (e.g., germy) onto Drawing Type (Violent = −0.5, Germy = 0.5), and we 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, we found that the Germy drawings appeared to have stronger infection-related features than less infection-related features (top left plot of Figure 3; see the supplemental repository for pairwise tests). In particular, the Germy drawings received high average germy, fatigue, and unhealthy ratings (i.e., reverse-scored healthy).

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 4 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. We 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, we 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.

Do people want to avoid germy persons more than violent persons? We 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), although work in this literature does not typically compare cues between threat categories as we 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, we recruited participants to complete a data-driven, 2-image forced choice reverse correlation task. In this task, participants chose hundreds of faces that they thought best represent a Germy person. We then averaged their choices 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 associ-
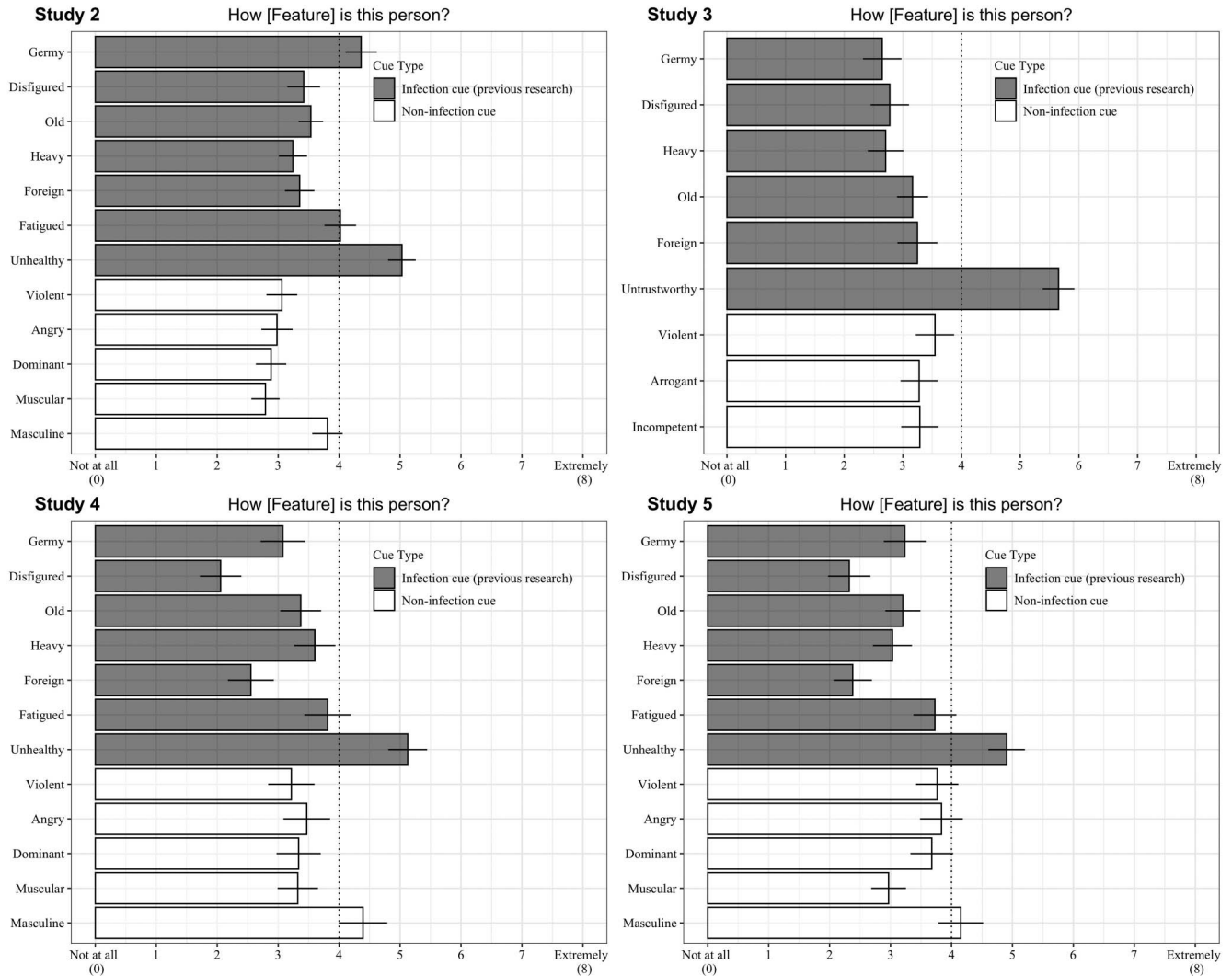
Figure 3. Mean feature ratings for the expectation-driven Germy representation (Study 2) and data-driven Germy and Infected representations (Studies 3–5). We colored the mean bars to highlight features that past research has categorized as cues associated with infection (gray fill) or has not examined in this context (white fill). We 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.

ated with infection and whether these features distinguish it from a noninfected person. Last, we assessed whether people want to avoid the person in the Germy mental image.

Method

Phase 1: Estimating mental images of a germy person.

Participants. We recruited undergraduate psychology student participants between February 26, 2016 and March 31, 2016 (see Table 1 for sample characteristics). We excluded two 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 two-image forced-choice reverse correlation image classification task, we first generated 400 pairs of stimuli using the rcicr package in R (Dotsch, 2016). Using the software, we 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 gray scale average of all male faces in the Karolinska Face Database (Lundqvist, Flykt, & Öhman, 1998; see Figure 5). 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
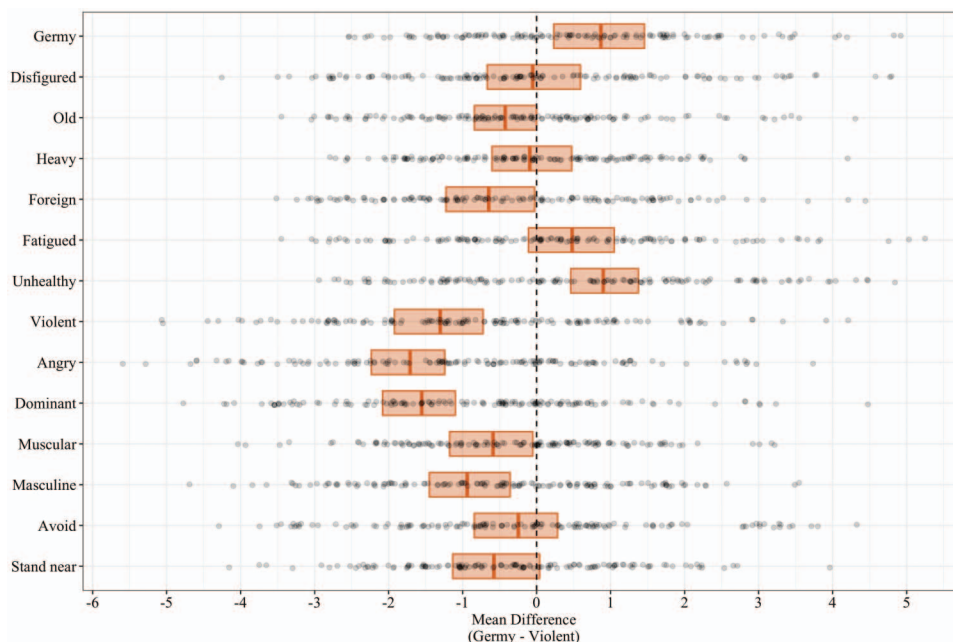
Figure 4. Mean differences in trait ratings between Germy and Violent drawings in Study 2. Dark, vertical lines inside crossbars (shaded boxes) depict fixed effects estimates for trait drawing condition (Germy–Violent). The widths of the crossbars represent 95% confidence intervals (bootstrap method, 1,000 resamples). Smaller, faded points depict observed difference scores (artist Germy drawing ratings minus their Violent drawing ratings). We also reverse-scored healthy (unhealthy). See the online article for the color version of this figure.

questionnaires that assessed trait-level threat concerns (see the supplemental repository).

Generating mental images: Classification images. Using rcicr (Dotsch, 2016), we 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 5). We did the same for the faces *not* chosen—the anticlassification 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. nonwelfare recipients (Brown-Iannuzzi et al., 2018)). See the non-Germy and Germy classification images in Figure 6.

Phase 2: Measuring subjective features of germy mental images.

Participants. We planned to recruit at least 260 participants using TurkPrime (Litman et al., 2017; see preregistration at https://

aspredicted.org/xp2ey.pdf). Between December 28, 2016 and December 29, 2016, we sampled participants until we had recruited approximately equal numbers between conditions and $N \geq 260$. Participants opened our survey 350 times, and we paid \$0.50 to all 290 who submitted their MTurk HIT assignments (82.86% submission rate). See Table 1 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. We 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 the supplemental repository). Following consent, participants used a 9-point scale to rate each face on clarity (i.e.,
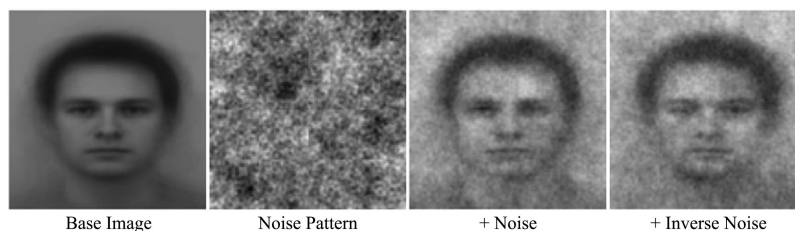

Base Image   Noise Pattern   + Noise   + Inverse Noise

Figure 5. 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). Male Neutral Straight (MNES) image from the Averaged Karolinska Directed Emotional Faces (AKDEF) database.
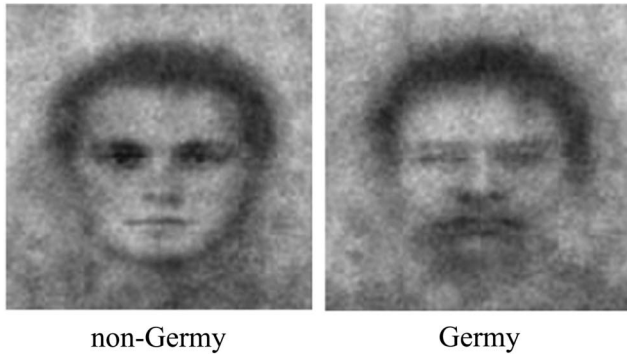
Figure 6. Images depict non-Germy and Germy classification images (Study 3). Male Neutral Straight (MNES) image from the Averaged Karolinska Directed Emotional Faces (AKDEF) database.

"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 rating portion of the study. Participants read definitions for nine features 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 features 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 features 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, we 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 features described above, we conducted a canonical discriminant analysis using the candisc function in the candisc package in R (Friendly & Fox, 2020), and we supplemented this analysis with univariate analyses. Specifically, we used the discriminant procedure to compute a linear combination of weights that—when applied to our observed feature 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 features—essentially a profile of features—rather than via each feature by itself

(which would ignore 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, Booth, & Irwing, 2012) as well as in implicit personality traits (Vianello, Schnabel, Sriram, & Nosek, 2013). To address interpretation issues due to heterogeneity, we used a heterogeneity coefficient—the equivalent proportion of variables coefficient (EPV)—to help estimate whether only one or a few features (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), we conducted Welch's *t* tests on the "want to avoid" and "willing to stand-near" items. In sum, we 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, we 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 3; see the supplemental repository for pairwise tests).

Do people discriminate between non-germy and germy mental images? Using canonical discriminant analyses, we found that the nine features combined well to maximally discriminate between the non-Germy and Germy mental images (cross-validated classification accuracy[2]: 78%, 95% CI [72%, 83%]), Canonical $R^2 = .28$, $F(9, 270) = 11.57$, $p < .001$ (see Figure 7A). With regard to specific features, 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 7B). Raters did not judge Germy representations to be significantly more arrogant than non-Germy representations ($d = -0.01$). When considering these feature differences and their correlations together, we observed Mahalanobis's $D = 1.25$, 95%

---

[2] 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 article. Higher AUROC values indicate that the model is better distinguishing a Germy face from a non-Germy face. We computed confidence intervals around AUROC via a bootstrapping procedure (2,000 stratified resamples).
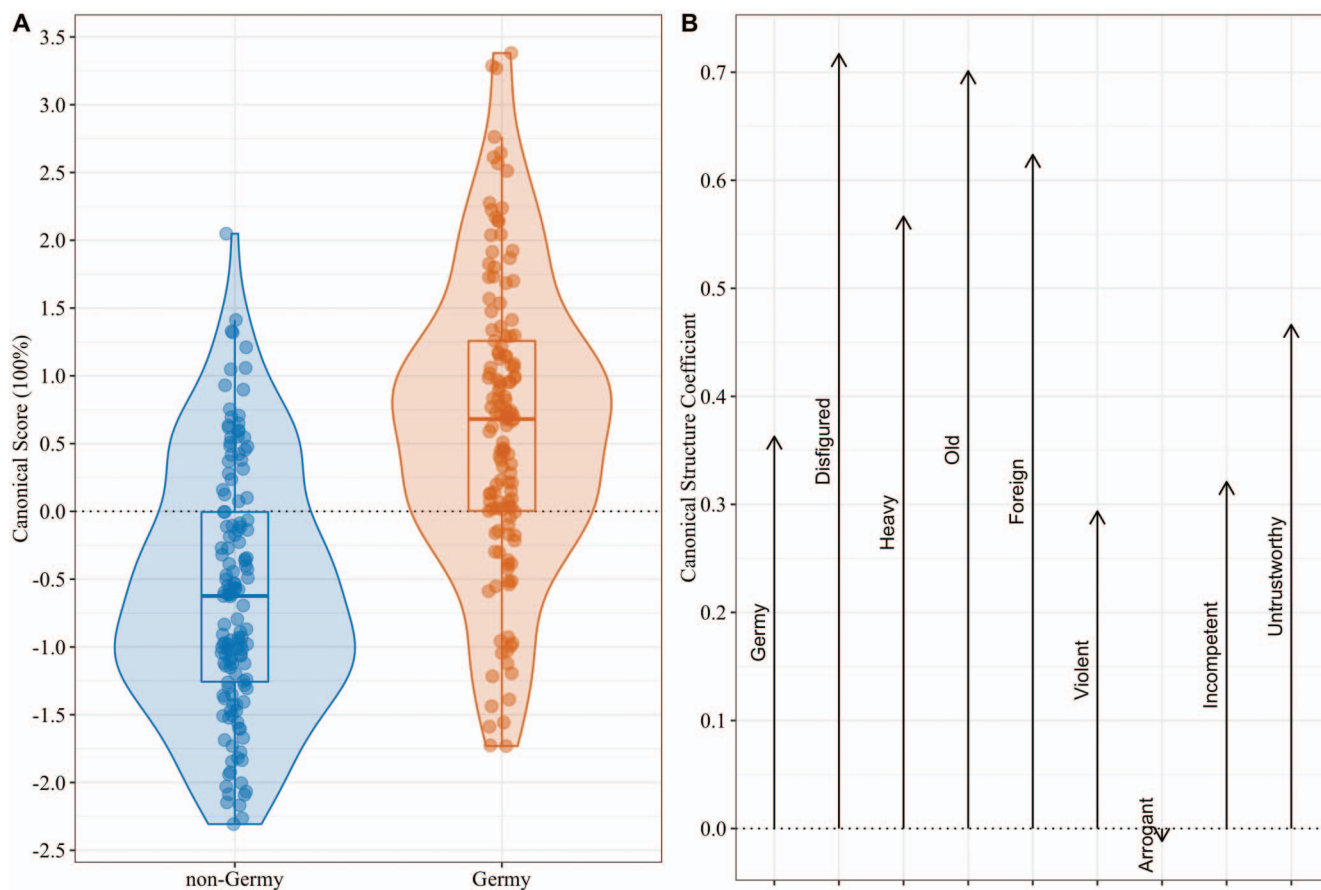
Figure 7. Together, the panels depict the maximal, configural difference between the non-Germy and Germy classification images (A) and the relative contribution of each rating to that difference (B; Study 3). The canonical scores represent participant ratings transformed to maximize the difference in 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 ratings and the canonical scores. Higher scores index the contribution of each feature to the non-Germy/Germy differences. See the online article for the color version of this figure.

CI [0.94, 1.42], which corresponds to 36.35% overlap (assuming multivariate normality). In other words, subjective 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, we observed some heterogeneity: The EPV suggests 51% (about 5) of the feature 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.

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 features. Examining only the Germy representation, we 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 with 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 anticlassification 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 anticlassification images can be psychologically meaningful (e.g., submissive representations appear similar to antidominant representations, (Dotsch & Todorov, 2012)). 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 choosing phase.

Given these possibilities, we made two key changes in Study 4. First, we replaced the Germy label with "Infected." Although 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), we 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. We recruited undergraduate psychology student participants between January 22, 2018 and March 12, 2018 (see preregistration at https://osf.io/9t4pr). We tabled pertinent sample characteristics in Table 1. We 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, we first generated 400 new pairs of stimuli (i.e., new random noise patterns) using the rcicr package in R (Dotsch, 2016). We used the same base face and procedure we 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: (a) Healthy, "having or showing good health; not sick or injured; the condition of being well or free from infectious disease"; or (b) 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, we created classification images and anticlassification images for participants' mental images of Infected and Healthy (see Figure 8). In total, we created one Infected classification image and one Healthy classification image as well as one Infected anticlassification image and one Healthy anticlassification image.

Phase 2: Measuring subjective features of infected and healthy mental images.

Participants. We planned to recruit 444 MTurkers using TurkPrime (Litman et al., 2017) so we might collect a final $N = 400$ if 10% of participants failed to meet inclusion criteria (see preregistration at https://osf.io/uy4dm). Between March 24 and 25, 2018, we sampled participants until we had recruited approximately equal numbers of MTurkers among four conditions and $N \geq 400$. Participants opened our survey 555 times and we paid $0.75 to all 464 (83.60% completion rate) participants who submitted their MTurk HIT assignments (see Table 1 for pertinent sample characteristics). We randomly assigned these participants
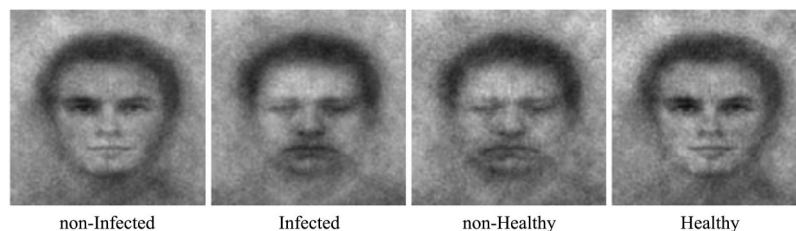

non-Infected   Infected   non-Healthy   Healthy

Figure 8. Images depict the non-Infected, Infected, non-Healthy, and Healthy classification images (Study 4). Male Neutral Straight (MNES) image from the Averaged Karolinska Directed Emotional Faces (AKDEF) database.

to rate either the Healthy anticlassification image ($n = 106$), the Healthy classification image ($n = 96$), the Infected anticlassification image ($n = 106$), or the Infected classification image ($n = 114$). As in Study 3, participants also rated additional images assessing individual difference effects (see the supplemental repository).

Statistical power. Our power analyses reflect our preregistration in which we 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 features that included more infectious disease threat and noninfectious disease threat features: 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, we 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 3; see the supplemental repository for pairwise tests).

Do people discriminate between healthy and infected mental images? Using canonical discriminant analyses, we found that the 12 features 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 9A). 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 9B). 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, we observed Mahalanobis's $D = 1.74$, 95% CI [1.30, 1.95], which corresponds to 23.85% overlap (assuming multivariate normality). Importantly, we observed some heterogeneity: The EPV suggests 45.48% (about 5) of the features 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.

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 wil-

ling 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, we 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 features 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 anticlassification 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. Although 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 per-
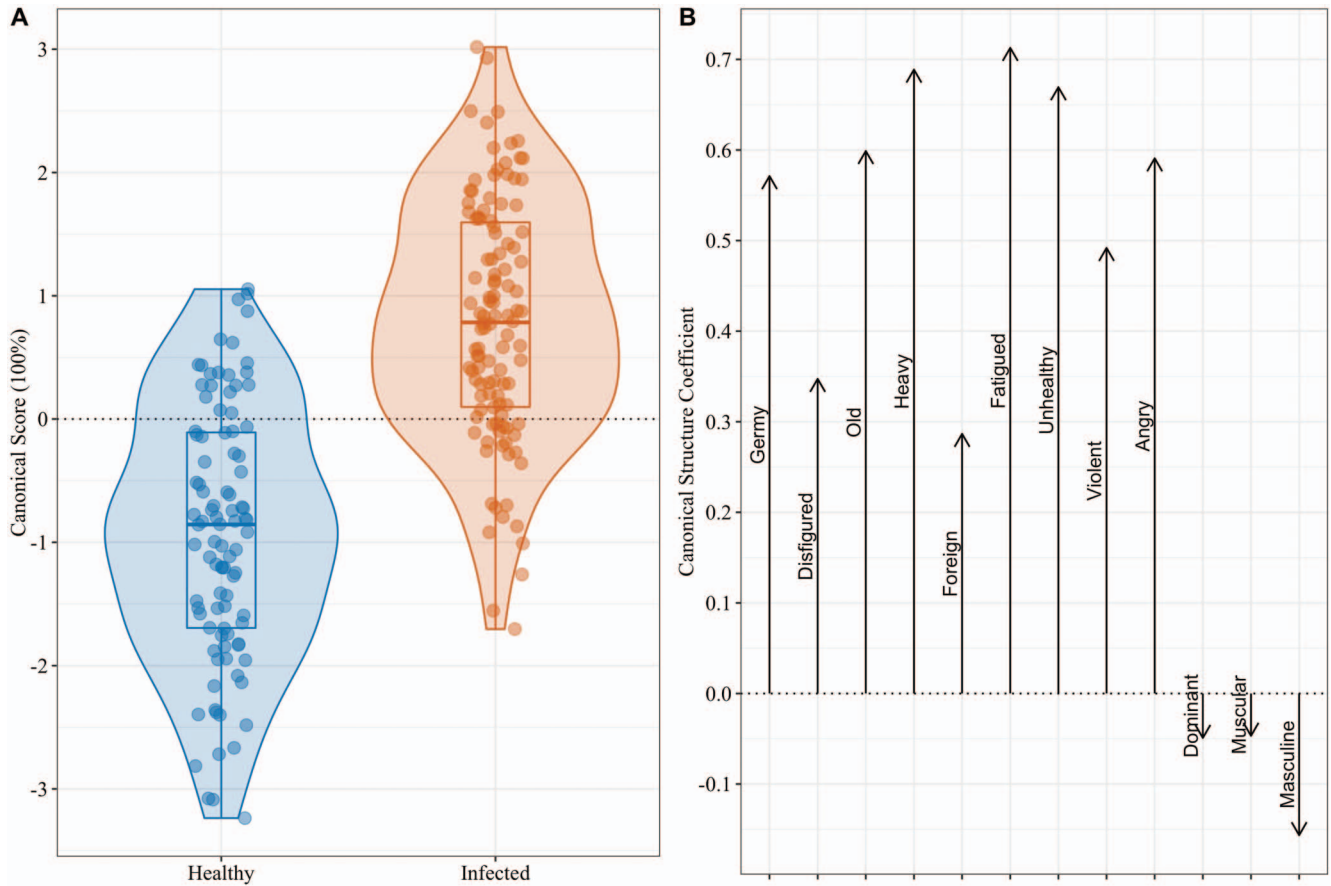
Figure 9. Together, the panels depict the maximal, configural difference between the Healthy and Infected classification images (A) and the relative contribution of each rating to that difference (B; Study 4). The canonical scores represent participant ratings transformed to maximize the difference in 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 ratings and the canonical scores. Higher scores index the contribution of each feature to the Healthy/Infected differences. See the online article for the color version of this figure.

sons. 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. We 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. We 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 9 and 10, 2017, we recruited participants until we had approximately equal numbers of participants between conditions

and $N \geq 200$. Participants opened our survey 398 times and we paid \$2.00 to all 244 (61.46% completion rate) participants who submitted their MTurk HIT assignments (see pertinent sample characteristics in Table 1). We 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.t-est function in the pwr package in R; (Champely, 2018)).

Procedure. For the reverse correlation task, we 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: (a) Violent, "prone to commit acts of violence; uses physical force intended to hurt, damage, or kill someone or something"; or (n) 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 choosing 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, we created classification images and anticlassification images for chooser participants' mental images of Violent and Germy (see Figure 10). In total, we created one Violent and one Germy classification image as well as one Violent anticlassification image and one Germy anticlassification image. As in Study 4, we only report analyses on classification images.

Phase 2: Measuring subjective features of violent and germy mental images.

Participants. We planned to recruit 500 MTurkers using TurkPrime (Litman et al., 2017; see preregistration at https://aspredicted.org/q6wv2.pdf). Between June 13 and 20, 2017, we sampled participants until we had recruited approximately equal numbers of MTurkers among four conditions and $N \geq 500$. Participants opened our survey 726 times and we 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 1 for pertinent sample characteristics). We randomly assigned these participants to rate a Violent anticlassification image ($n = 124$), a Violent classification image ($n = 127$), a Germy anticlassification 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 the supplemental repository).

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, we 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 3; see the supplemental repository for pairwise tests).

Do people discriminate between violent and germy mental images? Using canonical discriminant analyses, we found that the 12 subjective features 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 11A). 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 11B). 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, we observed Mahalanobis's $D = 1.05$, 95% CI [0.77, 1.20], which corresponds to 42.70% overlap (assuming multivariate normality). Importantly, we observed some heterogeneity: The EPV suggests 36.61% (about 4) of the features 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.

Do people want to avoid germy persons more than violent persons? We 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, we 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 we 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, we 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 features measured, perceptions of foreignness, violence, anger, and dominance contributed most strongly to configural differences between Violent and Germy
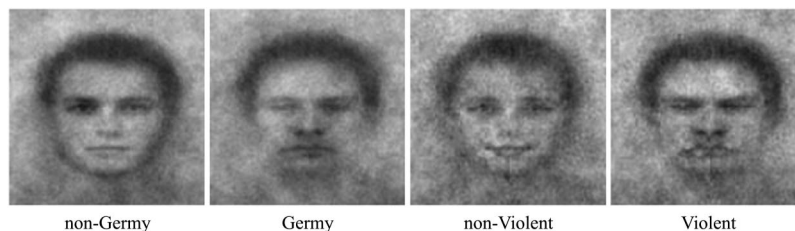

Figure 10. Images depict non-Germy (top left), Germy (top right), non-Violent (bottom left), and Violent (bottom right) classification images (Study 5). Male Neutral Straight (MNES) image from the Averaged Karolinska Directed Emotional Faces (AKDEF) database.
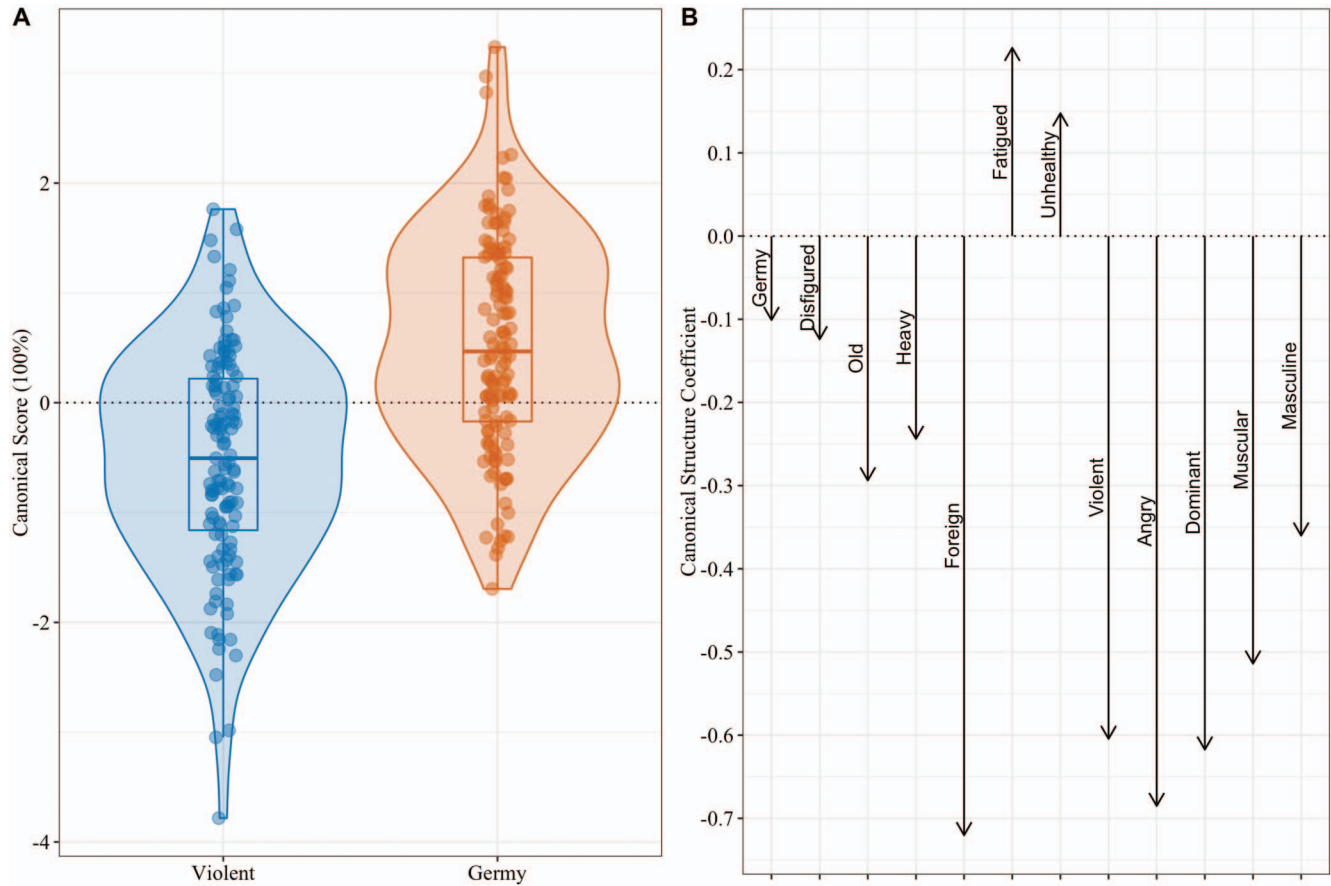
Figure 11. Together, the panels depict the maximal, configural difference between the Violent and Germy classification images (A) and the relative contribution of each rating to that difference (B; Study 5). The canonical scores represent participant ratings transformed to maximize the difference in 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 ratings and the canonical scores. Higher scores index the contribution of each feature to the Violent/Germy differences. See the online article for the color version of this figure.

mental images. These contrasts suggest Germy mental images—when compared with Violent mental images—appear to *share* many of the same disease threat-specific features 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,[3] 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, we 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 that people represent both Germy and Violent persons with features from multiple specific

---

[3] 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.

threats, but Violent representations possess more prominent, physical harm threat-specific features. We interpret these patterns as being most consistent with the threat-combination hypothesis.

General Discussion

We started this research with a simple question: How do people mentally represent distinct interpersonal threats? We proposed two hypotheses. The threat-specificity hypothesis predicts that people mentally represent distinct interpersonal threats with features specific to and diagnostic of 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, we 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. Specifically, 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: (a) 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 (b) 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, Petersen, & Arceneaux, 2017; Faulkner, Schaller, Park, & Duncan, 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 that we found that 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 we employed here have many strengths such as the fact that 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 choices 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 choices 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, we found it useful to compare representations with and without this constraint feature.

Second, 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. Though this averaging component limits some infection-relevant features from emerging in classification images, it does not limit all relevant features. 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 and 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. 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 rated equally). 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. We detailed a primary one earlier—expectation-driven methods allow participants to edit their representations, which may look very different compared with their spontaneous representation. In addition, expectation-driven methods are very 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 by 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. We are not aware of such a task, but we 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), we 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. We 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 grayscale base images are standard in the literature (Brinkman et al., 2017), 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 and colleagues (2015) are developing a task that incorporates color using a "Bubble Warp" approach, using random, colored image fragments rather than grayscale 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, we 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. We 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 multimethod approach may be our best option for fully capturing how people represent threat in their mind's eye.

Context Paragraph

We have been involved with a number of research projects within the conceptual space of pathogen threat and infectious disease avoidance. To date, most research (including our own) on such threats investigates piecemeal processes—such as perceiver reactions, evaluations, and associations—by presenting participants with targeted cues that theoretically indicate threat. The lead author recognized that theoretically motivated approaches like these may limit the choices that researchers make about the structure and content of their studies. To address this, we focused on methods that prioritize more spontaneous, participant-generated representations. We were surprised that different threat category representations shared many threatening features when they were generated via a reverse correlation paradigm but appeared more distinct when generated via more deliberative drawing and trait listing tasks. Different measurement methods lead to different interpretations. So, we think that future work on threat representations in particular and mental representations of social categories in general could benefit from approaching such representations like construct development. That is, researchers could consult experts and relevant participant populations on their opinions about target social categories (e.g., which features are important?), and they could use multiple tools to measure different aspects of mental representations. Ideally, well-developed mental representations of social categories would yield insight into psychological processes related to those categories.

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. *The American Political Science Review, 111,* 277–294. http://dx.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,* 478–485. http://dx.doi.org/10.1016/j.jesp.2008.12.008

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. http://dx.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,* 235–246. http://dx.doi.org/10.1037/0022-3514.53.2.235

Barrett, H. C. (2012). A hierarchical model of the evolution of human brain specializations. *Proceedings of the National Academy of Sciences, 109*(Suppl. 1), 10733–10740.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48.

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,* 333–361. http://dx.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,* 92–103. http://dx.doi.org/10.1177/0956797616674999

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,* 292–297. http://dx.doi.org/10.1037/xge0000376

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). New York, NY: Guilford Press.

Champely, S. (2018). pwr: Basic functions for power analysis (Version 1.2–2) [Computer software]. Retrieved from https://CRAN.R-project.org/package=pwr

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture,* (pp. 85–116). Cambridge, UK: Cambridge University Press.

Del Giudice, M. (2017). Heterogeneity coefficients for Mahalanobis' D as a multivariate effect size. *Multivariate Behavioral Research, 52,* 216–221. http://dx.doi.org/10.1080/00273171.2016.1262237

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

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18. http://dx.doi.org/10.1037/0022-3514.56.1.5

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

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3,* 562–571. http://dx.doi.org/10.1177/1948550611430272

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

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,* 97–115. http://dx.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,* 541–546. http://dx.doi.org/10.1016/j.paid.2009.05.001

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317. http://dx.doi.org/10.1037/0033-295X.95.3.307

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

Fellows, I. (2018). wordcloud: Word clouds (Version 2.6) [Computer software]. Retrieved from https://CRAN.R-project.org/package=wordcloud

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,* e32751. http://dx.doi.org/10.1371/journal.pone.0032751

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, 878–902. http://dx.doi.org/10.1037/0022-3514 .82.6.878

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

Friendly, M., & Fox, J. (2020). candisc: Visualizing generalized canonical discriminant and canonical correlation analysis (Version 0.8–3) [Computer software]. Retrieved from 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. http://dx.doi.org/10.1167/15.12.420

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

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223–233. http://dx.doi.org/10.1016/S1364-6613(00)01482-0

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). New York, NY: Psychology Press.

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,* 455–461. http://dx.doi.org/10.1037/a0035826

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14,* 640–643. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1478.x

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,* 601–625. http://dx.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. *The Journal of Neuroscience, 17,* 4302–4311. http://dx.doi.org/ 10.1523/JNEUROSCI.17-11-04302.1997

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,* 675–709. http://dx.doi.org/10.1037/ pspa0000046

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

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

Lovakov, A., & Agadullina, E. (2017). *Empirically derived guidelines for interpreting effect size in social psychology.* PsyArXiv. http://dx.doi.org/ 10.31234/osf.io/2epc4

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces* [Database of standardized facial images]. Stockholm, Sweden: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science, 28,* 209–226. http://dx.doi.org/10.1207/s15516709cog2802_4

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics, 5,* 269–287. http://dx .doi.org/10.3102/10769986005003269

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,* 1256–1265. http://dx.doi.org/10.1093/ cercor/bhh087

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

Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience and Biobehavioral Reviews, 35,* 1042–1051. http://dx.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,* 303–321. http://dx .doi.org/10.1037/a0014823

Park, J. H., Schaller, M., & Crandall, C. S. (2007). Pathogen-avoidance mechanisms and the stigmatization of obese people. *Evolution and Human Behavior, 28,* 410–414. http://dx.doi.org/10.1016/j.evolhumbehav.2007.05 .008

Queiroz, G. D., Hvitfeldt, E., Keyes, O., Misra, K., Mastny, T., Erickson, J., . . . Silge, J. (2019). tidytext: Text Mining using "dplyr", "ggplot2", and Other Tidy Tools (Version 0.2.2) [Computer software]. Retrieved from https://CRAN.R-project.org/package=tidytext

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,* 897–911. http://dx.doi.org/10 .1037/a0036498

R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. Retrieved from https://www.R-project .org/

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,* 77. http:// dx.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,* 639–646. http://dx.doi.org/10.1016/j.evolhumbehav.2012 .04.001

Schaller, M., Park, J., & Faulkner, J. (2003). Prehistoric dangers and contemporary prejudices. *European Review of Social Psychology, 14,* 105–137. http://dx.doi.org/10.1080/10463280340000036

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70,* 1126–1141. http://dx.doi.org/10.1037/0022-3514.70.6.1126

Stangor, C., & Lange, J. E. (1994). Mental representations of social groups: Advances in understanding stereotypes and stereotyping. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 357– 416). San Diego, CA: Elsevier. http://dx.doi.org/10.1016/S0065- 2601(08)60157-4

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,* 207–218. http://dx.doi.org/10.1037/ 0022-3514.62.2.207

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior, 27,* 131–144. http://dx.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,* 775–791. http://dx.doi.org/10.1111/j.1751-9004.2011.00389.x

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19– 136). New York, NY: Oxford University Press.

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 Series B, Biological Sciences, 366,* 3375–3388. http://dx.doi.org/10.1098/rstb.2011.0136

van Lunzen, J., & Altfeld, M. (2014). Sex Differences in Infectious Diseases–Common but Neglected. *The Journal of Infectious Diseases, 209*(Suppl. 3), S79–S80. http://dx.doi.org/10.1093/infdis/jiu159

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-0-387-21706-2

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences, 55,* 994–999. http://dx.doi.org/10.1016/j.paid.2013.08.008

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). New York, NY: Guilford Press.

Zebrowitz, L. A., & Franklin, R. G., Jr. (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,* 375–393. http://dx.doi.org/10.1080/0361073X.2014.897151