**"They Had It Coming!"**

**The Effect of Moral Character on Somatic and Mental Health Judgments**

*Somogy Varga, Andrew J. Latham, & Edouard Machery*

**Abstract**: Prior research has unveiled a *pathologization effect* where individuals perceived as having bad moral character are more likely to have their conditions labeled as diseases and are less often considered healthy compared to those viewed as having a good moral character. Moreover, these individuals are perceived as less unlucky in their affliction and more deserving of it. This study explores the broader impacts of moral character on such judgments, hypothesizing that these effects reach deeper and extend to both negative and positive moral evaluations. The pathologization effect also raises concerns about potential discrimination and the overmedicalization of normal health variations, so we also examine whether providing more detailed descriptions of conditions mitigates the influence of judgments of moral character. The methodology and broader implications of our findings are discussed, emphasizing the need for a deeper understanding of how moral judgments might influence patient care.

**Keywords**: Moral character, health, disease, experimental philosophy, pathologization

We typically assume that when we judge that a person is healthy or has a condition which is a disease, this judgment is based on certain descriptive criteria. However, a recent study suggests that the moral character of the person we are evaluating can influence these judgments (Varga, Latham & Machery, ms.). Individuals perceived as having a bad moral character were more likely to have their conditions judged to be diseases and be considered less healthy compared to those perceived as having a good moral character. This effect was called the *pathologization effect*. And this pathologization effect was hypothesized to be a consequence of people’s *just world beliefs*, as people perceived as having bad moral characters were also judged to be less unlucky regarding their condition and more deserving of it.

These findings raise several issues. First, there is a risk of discrimination against individuals *perceived* to have bad moral character, even if they do not. The judgments that these individuals are less healthy could lead to overmedicalization—where normal variations in health are treated as pathologies—and perhaps even overtreatment. These consequences underscore the importance of deepening our understanding of this effect and of exploring factors that could mitigate its impact. In the current study, we explore these issues through the lens of three overarching hypotheses. First, we aim to confirm that the pathologization effect is due to people perceived as having a bad character being judged more unhealthy and more likely to have a disease, in contrast to people perceived as having a good character who are judged more healthy, less likely to have a disease or both. We do this by also considering judgements about people who have an “ordinary” moral character.[[1]](#footnote-1) Second, we aim to examine whether the effect of moral character applies to both person-level and subpersonal judgments (e.g., judgments about specific organs or bodily systems). Finally, we examine whether providing information about the underlying mechanisms of a condition might mitigate the potentially problematic influence of moral character on judgments of health and disease.

We proceed as follows: In Section 1, we explore two pathways through which evaluative judgments may influence health and disease judgments, and identify open questions left by the previous study. After that, we will elaborate on this study’s hypotheses. Sections 2 and 3 detail our study's methodology and present our findings. Finally, in Section 4, we discuss the implications of our results, exploring how they contribute to the broader field and what they suggest for future research.

1. **Background and hypotheses**

Psychological and philosophical research offer insights into how evaluative judgments, including moral judgments, may influence health and disease judgments. Studies show that negative evaluative judgments and the impulse to blame significantly influence how other concepts are applied (Knobe, 2010), for example in assessments of causality (Hitchcock and Knobe, 2009) and intentional actions (Knobe, 2003). It has been suggested that disapproval of a person not only triggers a motivation to blame them but also shapes perceptions of their actions and character in ways that justify this blame (Alicke et al., 1994; Alicke, 2000). For example, this might involve attributing free will or intentional actions to them, the absence of which could excuse them and prevent assigning blame. If the impulse to blame significantly influences how some concepts are used, this influence may extend to the deployment of concepts of health and disease. Therefore, one might hypothesize that if people are inclined to blame someone based on their character or choices, they might hesitate to classify a condition of this individual as a disease, as labeling the condition as a disease could perhaps serve as an excusing factor, thereby obstructing the assignment of blame.[[2]](#footnote-2)

However, a recent study found evidence that perceived bad moral character can have the opposite effect: those perceived to have bad moral character can be more likely to have their conditions labeled as diseases and be considered less healthy compared to those perceived to have good moral character. Furthermore, these individuals are judged to be less unlucky in their affliction and more deserving of it, suggesting an influence of people’s moral evaluations (Varga, Latham & Machery, ms.). This effect was called the *pathologization effect*.

The pathologization effect is consistent with the Just-World Theory in psychology, according to which people assume that the world is fundamentally just in the sense that individuals generally get what they deserve (Lerner, 1977, 1980; for a review, see Furnham, 2003). As a result, participants are more inclined to classify the conditions of individuals perceived morally bad as diseases and view them as less healthy, due to the belief in a just-world framework where bad things, including diseases, are seen as fitting consequences for bad people. Conversely, good things, including health, may be seen as suitable consequences for good people. Importantly, just-world beliefs are taken to serve an adaptive function, making them difficult to abandon (Lerner and Miller, 1978), so much so that people will sometimes attribute blame to, for instance, accident victims, rape survivors, and patients with serious diseases or mental illnesses, rather than revise their fallacious belief (Montada, 1998; Braman & Lambert, 2001; Anderson, 1992; Bizer, Hart, & Jekogian, 2012).

These findings highlight several concerning issues. For instance, it seems likely that the pathologization effect could increase the risk of discrimination against individuals fairly or unfairly *perceived* to have bad moral character (e.g., racial or ethnic minorities). Additionally, if this effect is strong and extends to healthcare contexts, it might lead to overmedicalization and overtreatment, where normal variations in health are treated as medical conditions. Such outcomes underscore the need for further exploration, particularly in addressing in more detail the impact of moral character and identifying potential mitigating factors. In the sections that follow, we identify three critical issues left unresolved and describe how this study aims to address these gaps and contribute new insights.

**1.1. Negative vs. positive moral evaluation**

Varga, Latham & Machery (ms.) compared health and disease judgments when evaluating someone perceived to have a bad moral character and someone perceived to have a good moral character. And as described previously, they reported that someone perceived to have a bad moral character was more likely to be judged to be unhealthy and have a condition that is a disease: the pathologization effect. But, if people hold just-world beliefs, then it is plausible that positive moral evaluations also influence people’s health and disease judgments, so that people are more likely to be judged to be healthy and less likely to be judged to have a condition that is a disease if they are perceived to have a good moral character. Call this the *depathologization effect*. Indeed, the effects of moral praise and blame are often taken to be symmetrical in the philosophical literature (e.g., Shoemaker 2024). Examining whether there is a pathologization effect, depathologization effect, or both, on people’s health and disease judgments requires an appropriate baseline, so in this study we examine people’s judgments for someone who has an ordinary moral character.

**(Hypothesis 1)**: ***Depathologization***. Compared to individuals perceived as having evil or ordinary moral characters, there will be stronger agreement that moral saints are healthy and less agreement that they are afflicted with a disease, disorder, or dysfunction; additionally, there will be stronger agreement that the condition is harmful and unlucky, and stronger disagreement that it is deserved when the person is a moral saint.

If such a depathologization effect exists and is strong enough, then it raises concerns about the potential for health issues to be overlooked or minimized. Conditions that might actually be pathologies are treated as normal variations in health due to the perception of highly virtuous moral character, which could lead to undermedicalization and undertreatment. To investigate this, the current study examines the effects of moral character across three categories: moral saints, individuals with ordinary moral character, and those with evil moral character. The comparison between individuals with ordinary moral character and moral saints is particularly important, as it could reveal whether positive moral evaluations lead to a depathologization effect.

**1.2. The depth of the pathologization and depathologization effects**

Varga, Latham & Machery (ms.) primarily focused on person-level health and disease judgments (i.e., whether or not a person is judged to be healthy or has a disease), but did not explore subsystem-level judgments (i.e., whether an organ or subsystem of a person is healthy or has a disease). One might think that the influence of moral character judgments primarily affects person-level health assessments and might not extend deeply enough to influence evaluations of specific organs or subsystems. It is conceivable then that the influence of moral evaluations decreases or even disappears when the focus shifts to move away from personal attributes or behaviors.

However, given the previous results, one might hypothesize that the influence of moral evaluations is pervasive and extends to subpersonal judgments, such as evaluations of specific bodily or cognitive subsystems. This has led us to the second hypothesis.

**Hypothesis 2**: The influence of moral evaluations (that is the pathologization and depathologization effects) will extend to subpersonal judgments.

If these effects extend to subpersonal judgments and in particular if they extend to healthcare personnel, then this would indicate a more pervasive influence affecting even the more technical and specific aspects of medical care. To investigate this, the current study set out to examine the effects of moral character on both person-level health and disease judgments as well as subsystem-level judgments. This involved asking participants not only whether a person is healthy or has a disease but also whether specific subsystems implicated in the condition—the vascular system and the cognitive-emotional subsystem—are considered healthy or diseased.

**1.3. Levels of description and potential mitigating effects**

We also sought to determine if presenting only the symptoms of a condition versus adding more detail about the mechanisms of a condition would influence health-related judgments. For this, we presented participants with two types of cases: one in which they made judgments based solely on a symptom-only description of the condition, and another where these symptoms were contextualized and explained in terms of underlying mechanisms. We predicted that providing more detailed information about the underlying mechanisms might mitigate the influence of moral character judgments on both person-level and subsystem-level judgments. More precisely:

**Hypothesis 3**: In symptom-only description cases, people will agree more with the claim that the condition is healthy and disagree more with the claim that it is a disease, disorder, or dysfunction. Also, there will be stronger disagreement that the condition is harmful, unlucky, and stronger agreement that it is deserved.

In mechanism-description cases, people will disagree more with the claim that the condition is healthy and agree more with the claim that it is a disease, disorder, or dysfunction. Also, there will be stronger agreement that the condition is harmful and unlucky, and stronger disagreement that it is deserved.

Varga, Latham and Machery (ms.) reported a pathologization effect but did not explore potential mitigating factors. Adding information about the causal mechanisms beyond mere symptoms might mitigate the impact of moral character on health-related judgments. Two main reasons support this idea. First, detailed knowledge about the underlying mechanisms of a condition might help anchor judgments in the description of those mechanisms. When participants understand the underlying physiological or cognitive mechanisms of a condition, their assessments are more likely to be based on these mechanisms than on (assumptions) about an individual's character. Second, moral judgments often fill gaps where there is uncertainty or a lack of understanding about a condition. For instance, if one is uncertain whether a condition correctly counts as a disease or not, people might sort conditions according to, for instance, their standing Just-World beliefs. This means that a condition might count as a disease if the person is perceived to have a negative moral character (the pathologization effect), but the very same condition possessed by someone perceived to have a good moral character might count as normal variation (the depathologization effect). By clearly explaining how and why a condition occurs, explanations can reduce the ambiguity that might otherwise lead moral evaluators to fill knowledge gaps.

Exploring potential mitigating factors is crucial for several reasons. Identifying mitigating factors is key to reducing the risk of discrimination based on moral character assessments and enhancing fairness within healthcare settings. But it can also provide valuable insights for addressing similar biases in non-medical settings like employment. If an employee is perceived as less healthy due to negative judgments about their moral character, then this might lead to assumptions about, for example, their capacity to cope with stress or meet physical demands, and absenteeism or productivity. Based on such assumptions, managers might doubt the employee's capabilities, leading them to overlook them for promotions and other opportunities.

1. **Experimental Design**

The study was pre-registered at OSF [osf.io/3zx6d]. We used G\*Power to conduct a power analysis. Our goal was to obtain 0.9 power to detect a medium effect size at .005 alpha error probability (Benjamin et al. 2018). We calculated that a sample size of 348 participants would be required. To be safe, we increased the sample by 15% to account for attention and comprehension check failures. This gave us a total sample size of 400. We recruited 400 native English speakers online using Prolific. 13 were excluded from the analyses for failing to respond to all the questions or answer all the attention and comprehension checks correctly. The final sample consisted of 387 participants (187 female; 14 Trans/Non-Binary; 186 male, age 18-79, M=36.69; SD=11.55). Ethics approval for the study was obtained from the Aarhus University Human Ethics Committee.

The study employs the contrastive vignette technique to test what factors influence people’s judgments. The vignettes include experimental elements that are manipulated across scenarios and controlled elements that remain constant across scenarios. The design facilitates tracking the impact of manipulated factors by systematically contrasting participants' judgments across various scenarios, aiming to determine how closely these judgments align with the hypotheses.

The study assessed the influence of three distinct factors on participants' judgments. The first factor, "*character*," manipulates the moral character of the main person in the scenario, Alex, and has three levels, evil, ordinary, and moral saint. The second factor, "*condition nature*," alters the type of medical ailment Alex faces and has two levels. Alex's condition either impacts his sexual capabilities, primarily his ability to maintain an erection (erection), or manifests as disruptive anxiety affecting his focus and attention (anxiety). The third factor, "*level of description*," modifies how Alex's condition is presented and has two levels: it is either described solely in terms of symptoms or in a more detailed manner, with symptoms explained in relation to the underlying mechanisms or subsystems that are affected.

The study thus employs a 3 (character: evil, ordinary, saint) x 2 (condition nature: erection problems vs. anxiety) x 2 (symptom description vs. symptom and mechanism description) between-subjects design. This resulted in a total of twelve different scenarios, with participants randomly assigned to one of the twelve possible vignettes, each reflecting a unique combination of moral character, disease, and level of description. Following the vignette, participants were asked their level of agreement with six assertions, and the degree to which they judge that the condition described is dysfunctional and harmful. Participants were also asked to respond to three comprehension and manipulation check questions. The order of questions was fully randomized.

For demonstrative purposes, here is a vignette in which Alex has a saint-like character, the condition involves erectile problems, and there is an added explanation of the underlying subsystem:

Alex is a 45-year-old man with a saint-like moral character. He has committed his life to eradicating illiteracy and poverty in impoverished areas. Going far beyond what could be morally required of anyone, he relinquished his comfortable lifestyle in order to live modestly within the community that he serves, witnessing their daily hardships. Alex’s extraordinary self-sacrifice and unwavering dedication to improving the lives of others have earned him the profound admiration of all who know him. Lately, Alex has been experiencing trouble. Specifically, he has been experiencing a consistent inability to achieve or maintain an erection sufficient for sexual activity. Alex's vascular system is not functioning properly, impairing blood flow necessary for an erection.

Here is the list of the 12 possible scenarios:

|  |  |  |  |
| --- | --- | --- | --- |
| **Scenario** | **Character** | **Condition** | **Description** |
| **1** | Saint | Erection | No Description |
| **2** | Saint | Erection | Description |
| **3** | Saint | Anxiety | No Description |
| **4** | Saint | Anxiety | Description |
| **5** | Ordinary | Erection | No Description |
| **6** | Ordinary | Erection | Description |
| **7** | Ordinary | Anxiety | No Description |
| **8** | Ordinary | Anxiety | Description |
| **9** | Evil | Erection | No Description |
| **10** | Evil | Erection | Description |
| **11** | Evil | Anxiety | No Description |
| **12** | Evil | Anxiety | Description |

While the focus of our study was the influence of perceived character on judgments of health, disorder, and medicalization, we asked a broad range of questions.

First, to examine participant’s health and disease judgments we asked their level of agreement on a 7-point Likert scale with the statements "In this scenario, Alex is healthy," "In this scenario, Alex has a disease/disorder",, "In this scenario, Alex’s [Sub-Person System] is healthy," and "In this scenario, Alex’s [Sub-Person System] has a disease or disorder." The Likert scale ranged from "strongly disagree" to "strongly agree."

Second, participants were asked how harmful they perceived the condition to be: "On a scale ranging from 1 (not at all harmful) to 7 (very harmful), where would you place the condition described in this scenario?". Also, participants were asked to indicate their perceived level of dysfunction ("On a scale ranging from 1 (not at all a dysfunction) to 7 (very dysfunctional), where would you place the condition described in this scenario?").

Third, to examine participants’ evaluative judgments, rather than ask participants straightforwardly whether Alex was to blame for the condition, we examined participants’ judgments about several statements describing factors strongly associated with blame. These were: "In this scenario, Alex is unlucky to have developed the condition," and "In this scenario, Alex deserves to have the condition." As with the health and disease judgments, participants indicated their level of agreement on a 7-point Likert scale. Finally, participants were asked two attention check and comprehension questions: "In this scenario, how old is Alex?" and "In this scenario, what kind of moral character does Alex have?"

1. **Results**

We first examined participants’ judgements across conditions using a MANOVA.[[3]](#footnote-3) Gender, age, ethnicity, and political ideology were entered into the analysis as covariates. These covariates were entered into each of the analyses we report which follow. The results of the analysis revealed a significant main effect of moral character, Λ = .338, F(16, 724) = 32.624, p < .001, condition type, Λ = .826, F(8, 362) = 9.562, p < .001, level of description, Λ = .832, F(8, 362) = 9.126, p < .001. There was also a significant effect of the interaction between condition type and level of description, Λ = .942, F(8, 362) = 2.793, p = .005. We also observed a significant influence of the covariates ethnicity, Λ = .928, F(8, 362) = 3.512, p < .001, and political ideology, Λ = .932, F(8, 362) = 3.291, p = .001. Next, we report results of separate ANOVAs that show the effects of these factors on participants’ judgments (adjusting for multiple comparisons).

First, we examined the effects of moral character, condition type, and level of description on people’s health judgments. There was a significant main effect of moral character, F(2, 369) = 37.963, p < .001, and level of description, F(1, 369) = 12.937, p < .001. Pairwise comparisons with Bonferroni correction showed that people’s health judgments for evil moral character (M = 2.51, SD = 1.39) were significantly lower than for saint-like (M = 3.77, SD = 1.39; p < .001) and ordinary moral characters (M = 3.85, SD = 1.39; p < .001). There was no evidence of a difference in health judgments between saint-like and ordinary moral character (p > .999). The main effect of level of description was that people’s health judgments for symptom-only conditions (M = 3.63, SD = 1.39) were significantly higher than for mechanism conditions (M = 3.12, SD = 1.39).

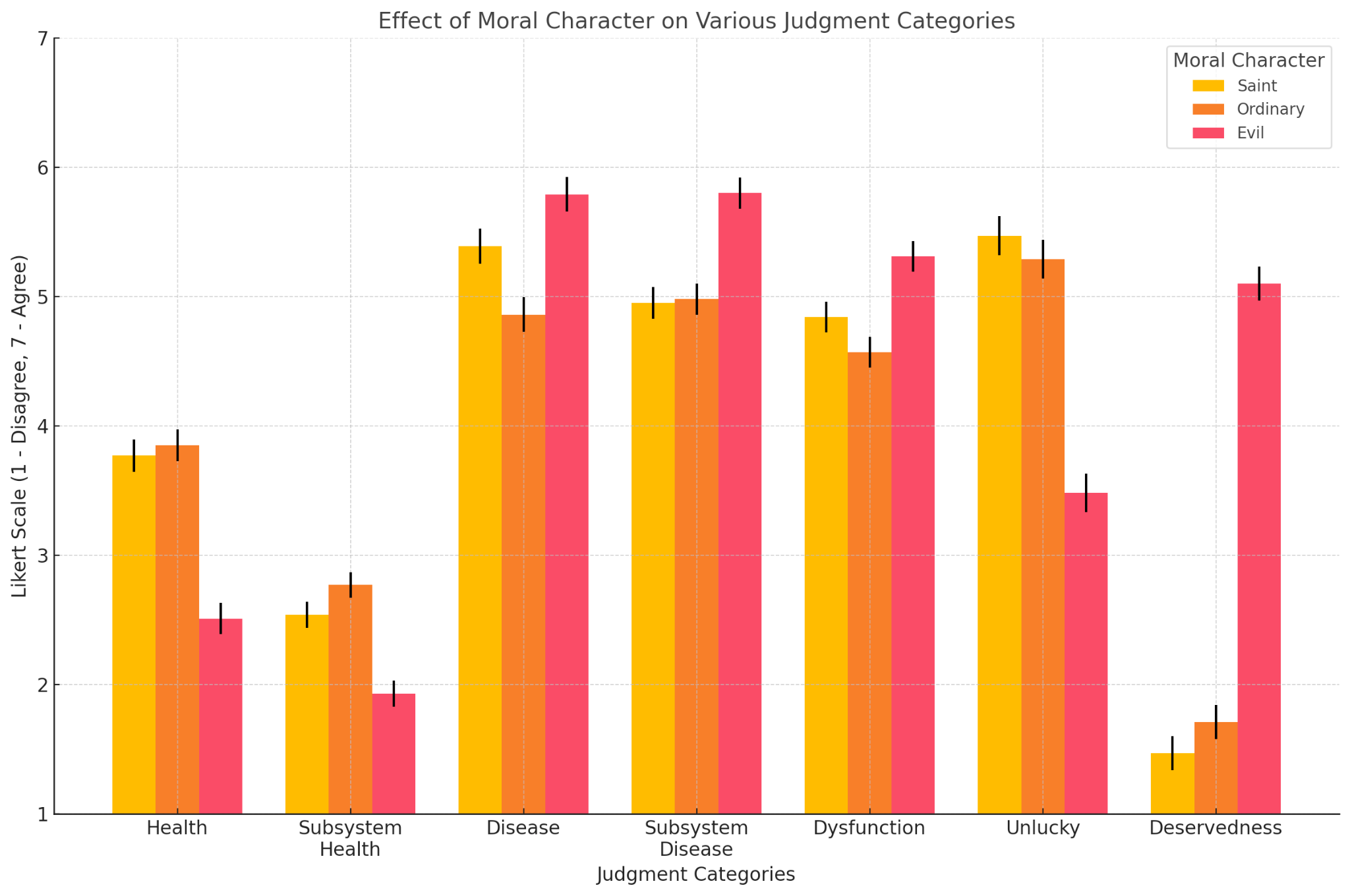
Next, we examined people’s subsystem health judgments. There was a significant main effect of moral character, F(2, 369) = 18.743, p < .001, and level of description, F(1, 369) = 54.082, p < .001. Once again, pairwise comparisons with Bonferroni correction showed that people’s health judgments for evil moral character (M = 1.93, SD = 1.14) were significantly lower than for saint-like (M = 2.54, SD = 1.14; p < .001) and ordinary moral characters (M = 2.77, SD = 1.14; p < .001). There was no evidence of a difference in health judgments between saint-like and ordinary moral characters (p = .318). The main effect of level of description was that people’s health judgments for symptom-only conditions (M = 2.84, SD = 1.13) were significantly higher than for mechanism conditions (M = 1.99, SD = 1.39). We were interested in exploring whether there was any difference between people’s person-level and sub-system level health judgments. A paired-sample t-test found that people’s person-level health judgments (M = 3.37, SD = 1.55), overall, were significantly higher than their sub-system level health judgments (M = 2.41, SD = 1.28; *t*(384) = 12.932, p < .001).

Second, we examined people’s disease judgments. There was only a significant main effect of moral character, F(2, 369) = 11.951, p < .001. Pairwise comparisons with Bonferroni correction showed that people’s disease judgments for evil moral character (M = 5.79, SD = 1.52) were significantly higher than for ordinary moral characters (M = 4.86, SD = 1.53; p < .001. However, there was no evidence of a difference in people’s disease judgments for saint-like moral character (M = 5.39, SD = 1.53) and both evil (p = .117) and ordinary moral characters (p = .018).

For people’s subsystem disease judgments there was a significant main effect of moral character, F(2, 369) = 16.156, p < .001, and level of description, F(1, 369) = 46.057, p < .001. Pairwise comparisons with Bonferroni correction showed that people’s subsystem disease judgments to evil moral characters (M = 5.80, SD = 1.37) were significantly higher than for saint-like (M = 4.95, SD = 1.38; p < .001) and ordinary moral characters (M = 4.98, SD = 1.38; p < .001). There was no evidence of a difference in subsystem disease judgments between saint-like and ordinary moral characters (p > .999). The main effect of level of description was that people’s subsystem disease judgments for symptom-only conditions (M = 4.77, SD = 1.37) were significantly lower than for mechanism conditions (M = 5.72, SD = 1.37).

Once again, we were interested in exploring whether there was any difference between people’s person-level and sub-system level judgments, this time for disease. A paired-sample t-test found no evidence of a difference between people’s person level disease judgments (M = 5.35, SD = 1.58) and sub-system level disease judgments (M = 5.25, SD = 1.51; *t*(384) = 1.283, p = .200).

Third, we examined people’s dysfunction judgments. There was only a significant main effect of moral character, F(2, 369) = 19.094, p < .001. Pairwise comparisons with Bonferroni correction showed that people’s dysfunction judgments for evil moral characters (M = 5.31, SD = 1.34) were significantly higher than for ordinary moral characters (M = 4.57, SD = 1.35; p < .001. There was no evidence of a difference in people’s dysfunction judgments for saint-like moral characters (M = 4.84, SD = 1.34) and both evil (p = .015) and ordinary moral characters (p = .349).



Fourth, we examined people’s harm judgments. There was a significant main effect of condition type, F(1, 369) = 25.776, p < .001, and interaction effect between moral character and condition type, F(2, 369) = 11.838, p < .001. There was also a significant effect of the covariate ethnicity, F(2, 369) = 18.676, p < .001. The main effect of condition type was that erectile dysfunction (M = 3.65, SD = 1.45) was judged to be significantly less harmful than anxiety (M = 4.41, SD = 1.45). Simple effects tests with Bonferroni correction were carried out on the two-way interaction between moral character and condition type. Firstly, in anxiety conditions, anxiety was judged to be more harmful to people with an evil moral character (M = 4.91, SD = 1.44) than an ordinary moral character (M = 3.89, SD = 1.45; p < .001). There was no evidence of a difference in people’s anxiety harm judgments to saint-like moral characters (M = 4.42, SD = 1.45) and both evil (p = .162) and ordinary moral characters (p = .105). Secondly, for erectile dysfunction, there was no evidence of a difference in people’s harm judgments between saint-like (M = 3.69, SD = 1.45), ordinary (M = 4.00, SD = 1.46) and evil moral characters (M = 3.26, SD = 1.46). Thirdly, for both saint-like and ordinary moral characters, there was no difference in people’s harm judgments between anxiety and erectile dysfunction (Saint-like: p = .005; Ordinary: p = .662). In contrast, for evil moral characters, anxiety was judged by people to be significantly more harmful than anxiety (p < .001).

Next, we examined people’s unlucky judgments (i.e., judgments about how unlucky the person was to develop the condition). There was only a significant main effect of moral character, F(2, 369) = 54.175, p < .001. Bonferroni correction showed that people’s unlucky judgments for evil moral characters (M = 3.48, SD = 1.69) were significantly lower than for saint-like (M = 5.47, SD = 1.71; p < .001) and ordinary moral characters (M = 5.29, SD = 1.70; p < .001). There was no evidence of a difference in unlucky judgments between saint-like and ordinary moral characters (p > .999).

Finally, we examined people’s deservedness judgments. Once again, there was only a significant main effect of moral character, F(2, 369) = 244.199, p < .001. Bonferroni correction showed that people’s deservedness judgments for evil moral characters (M = 5.10, SD = 1.48) were significantly higher than for saint-like (M = 1.47, SD = 1.48; p < .001) and ordinary moral characters (M = 1.71, SD = 1.48; p < .001). There was no evidence of a difference in unlucky judgments between saint-like and ordinary moral characters (p = .566).

1. **The Moral Psychology of Diagnosis**

Our study confirms the presence of a pathologization effect: individuals perceived as having evil moral character are more likely to have their conditions labeled as diseases and dysfunctions, and are less frequently considered healthy compared to those with ordinary moral character. Bad individuals are also viewed as less unfortunate in their affliction and as more deserving of it, once again suggestive of the influence of moral judgments on health assessments. These results are consistent with the Just-World Theory, which suggests that people tend to believe in a fundamentally just world where individuals generally receive the outcomes they deserve.

While we found further evidence for the pathologization effect, we found no evidence for the depathologization effect (Hypothesis one). We initially hypothesized that just-world beliefs might lead people to judge that moral saints are healthier, and that their conditions are less likely to be judged to be diseases or dysfunctions. We also predicted that people would judge that their conditions are more harmful and that they are less fortunate to have them, and that they are more undeserved. However, this hypothesis was not borne out by our results. Instead, both for person-level and subsystem-level health judgments, we found no evidence of a difference between saint-like and ordinary moral characters. Similarly, when it came to judgments of harm, how unlucky the person was to develop the condition, or how much they deserved it, there was no difference between saint-like and ordinary individuals. It appears there is no straightforward health bonus for moral saints compared to those with ordinary moral characters: their virtuous character does not lead to more favorable health assessments.

Of course, it might be that just-world beliefs do not result in moral saints being judged to be inherently healthier, nevertheless they might convey a ‘health bonus’ in other ways. For instance, being morally good might result in health traits being judged to be healthier. In the same way, morally evil people might not be judged to be inherently unhealthy, but instead traits already judged to be unhealthy might be judged to be unhealthier. The just-world framework might act as a heuristic that sorts negative health traits with negative moral character, and positive health traits with positive moral character, so that those positive and negative sorted health traits are evaluated more strongly than they otherwise would be. If that is right, then we would not predict a depathologization effect, at least of the sort hypothesized in this study, as we only examined the effect of moral character on negative health conditions. Rather, we should predict a health bonus of good moral character when examining positive health traits. Importantly, we should also *not* see a straightforward pathologization effect on positive health traits (in the same way that we do not see a depathologization effect on negative health traits in this study). This would be an interesting avenue for future research.

The findings from our study, where the pathologization effect was confirmed but the depathologization effect (at least as hypothesized) was not, highlight a possible asymmetry in the influence of just-world beliefs on health-related judgments. While the Just World Theory suggests a symmetrical relationship between moral character and health outcomes—where good things happen to good people and bad things to bad people—the psychological impact of a *negativity bias* might lead to an asymmetrical influence. This bias occurs because negative stimuli have a more profound psychological impact than positive stimuli, a well-established phenomenon in psychological research. This bias toward focusing on negative information, which is believed to serve vital evolutionary adaptive functions, ensures that negative information is processed more extensively and exerts a greater influence on psychological processes than neutral or positive information (see, e.g., Baumeister et al., 2001; Peeters & Czapinski, 1990). The negativity bias is particularly evident in the domain of morality, where research shows that people's moral perceptions are significantly more affected by negative traits than positive ones, and moral cognition is inherently more sensitive to assigning blame than offering praise (see Rozin & Royzman, 2001; Guglielmo & Malle, 2019). For example, when identical decisions lead to negative versus positive outcomes, individuals are more likely to assign blame for negative outcomes than they are to offer praise for positive ones (Guglielmo & Malle, 2019; Siegel et al., 2017).

The notion of negativity bias might help explain why the findings did not support the depathologization hypothesis. Negative moral evaluations and blame, being more salient and eliciting stronger emotional and cognitive responses than positive evaluations, lead to lower health judgments (pathologization) and perceptions of luck, but greater deservedness, whereas positive moral judgments and praise do not correspondingly enhance health judgments (depathologization) or perceptions of luck and deservedness. Overall, this might explain why saints do not receive a "health bonus" compared to individuals with ordinary moral character.[[4]](#footnote-4)

In addition, some of the results in disease and dysfunction judgments were puzzling. In judgments of subsystem-level diseases, saints were assessed similarly to individuals with ordinary moral character. However, the results in judgments of dysfunction are more unexpected, showing that saints' judgments straddle those of ordinary and evil characters, with no significant differences from either group after Bonferroni correction. Most surprisingly, in person-level disease assessments, saints fare no better than individuals perceived as morally evil, a puzzling outcome. Prior research shows that health and disease judgments come apart: a range of factors have differential effects on health judgments on the one hand and disease and dysfunction judgments on the other (Varga & Latham 2024a; Varga & Latham 2024b). Thus, participants seem to conceptualize health not merely as the absence of disease or dysfunction, but more holistically, as encompassing some additional qualities, perhaps like well-being, robustness, or vitality. However, while this helps us understand why health and disease judgments might diverge, it does not explain the specific pattern observed. But before exploring why this discrepancy occurs, further research should confirm whether these individual results are genuine, and not a mere result of sampling variation.

Hypothesis two proposed that the influence of moral character on health judgments would be uniform, regardless of whether the focus was on overall person-level health or more specific subsystem-level health assessments. We found support for this hypothesis suggesting a consistent influence of moral character across both broad and specific medical judgments: moral evaluations do not just affect general perceptions of whether someone is healthy or not, but also influence the assessments of their bodily systems and organs. Such pervasive influence of moral character gives rise to concerns about whether this effect, if found in health professionals, can affect patient care outcomes. Patients perceived as having a worse moral character could be at a disadvantage, which could affect their treatment outcomes and overall healthcare experience.

Our results with respect to Hypothesis 3 were mixed. Aligning somewhat with the patterns observed in the first hypothesis, the level of description differentially affected health and disease judgments. Specifically, judgments of health and subsystem health for symptom-only conditions were significantly higher than those for mechanism conditions, and subsystem disease judgments for symptom-only conditions were significantly lower than for mechanism conditions. This aspect of the findings can be roughly explained as follows: the absence of detailed information about the affected subsystem means participants have less specific data to inform their evaluations, which could result in less severe assessments of the condition. At the same time, the mechanism condition provides more relevant information, enabling people to be more certain in their judgments.

Aligned with prior findings that different factors impact health judgments and disease and dysfunction judgments, there were no significant effects of the level of description observed on person-level disease and dysfunction judgments in our study. One possible explanation is that the symptoms associated with these conditions are very clear and distinct, which might minimize the impact of additional explanatory details on judgments. Once a condition is clearly identified as a disease or dysfunction from symptoms, further mechanistic details may not significantly alter the perceived severity of the assessment. The limited effect of adding mechanistic information is underscored by the unchanged judgments about harm, unluckiness, and deservedness, despite the addition of such details.

Hypothesis 3 was also significant due to its potential to reveal something about whether additional description might mitigate the effect of perceived moral character. Providing detailed information about mechanisms beyond mere symptoms could reduce ambiguity and anchor judgments more firmly in the description of these mechanisms, which, in turn, could reduce the influence of perceived moral character. However, our findings do not support this. If adding mechanistic explanations to symptoms were effective in mitigating the influence of moral character on health-related judgments, an interaction effect between the level of description and moral character would be expected. However, no such interaction was observed, suggesting that detailed mechanistic explanations do not significantly alter the influence of moral character on health judgments compared to when only symptoms are presented. This highlights the necessity for strategies that go beyond providing mechanistic explanations, emphasizing the need for approaches that actively engage with and potentially reshape existing perceptions and biases.

**5. Conceptual Engineering of Health Concepts**

In recent years, many philosophers of medicine have conceptualized their goals anew: instead of aiming at *analyzing* the concepts of health and disease (either lay people’s or medical practitioners’), many now think that they should put forward *revisions* of these concepts (e.g., Schwartz, 2007; Gagné-Julien, 2024). In the jargon of contemporary metaphilosophy, they are engaged in *engineering* the concepts of health and disease (on conceptual engineering, see Haslanger 2000; Cappelen, 2018; Machery, 2017).

As a metaphilosophical framework for understanding philosophy, conceptual engineering has been the object of sustained discussion. Most relevant for our project is the *implementation question*: what are the practical challenges in actually changing people’s concepts? While these challenges are multifarious, an important kind of challenge is psychological in nature: our psychology might make implementing desirable conceptual changes to our concepts difficult (Fisher, 2020; Machery, in press).

Our research bears on conceptual engineering in two distinct ways. First, and perhaps more importantly, it justifies engineering the deployment of the concept of disease. (We refer to the “deployment of the concept of disease” rather than the “concept of disease” to remain neutral about whether the causal influence observed in our study is constitutive of the concept of disease: see Machery, 2008 for a discussion of the difficulty of answering this kind of question.) Philosophers of medicine disagree extensively about how to understand health and disease (see Murphy, 2021; Barnes 2023, Ch 1), but the following point at least would elicit, we suspect, a large consensus: whether a patient is a saint or a sinner should not matter to how sick or healthy they are. Naturalists such as Boorse, who hold that whether one is sick or healthy only depends on natural facts, and not on anyone’s values, would of course agree with this claim. Normativists, who, in contrast, hold that whether one is sick or healthy do depend on values, usually limit the role of evaluation to assessing the condition itself, not the broader moral character of the agent. Our findings suggest however that whether a patient is a sinner (but not whether they are a saint) influences disease judgments, as we just explained (see also Varga et al., ms). This empirical fact about the deployment of the concept of disease justifies engineering the concept of disease: how we use this concept should be reformed to remove this pernicious influence (but see next section on the limitations of our work). More generally, this proposal illustrates how conceptual engineering depends on understanding how concepts are actually used, and that empirical, including experimental, research is useful and maybe even necessary for such a project. The details of such use can be, our study illustrates, subtle and not intuitive at all: it was not antecedently obvious that moral judgments would influence diagnosis. (And indeed this influence is very specific, as is shown by the lack of evidence for our second hypothesis, that saints would be judged to be healthier.)

Secondly, the discussion of the implementation challenge has highlighted different ways in which a conceptual engineering proposal could be hampered by features of concepts. For instance, some inferences can be automatic and implicit, and as a result hard to suppress (Fischer, 2020); some concepts can be “attractors,” that is, override ways of thinking about their extension (Machery, in press; but see Köiv, 2024 for discussion). Our findings identify yet another way in which implementing an engineering proposal can be challenging: intuitive ways of remedying how a concept–in this case the concept of disease–is deployed might turn out to be inefficient. It would have been natural to try to undercut the pernicious influence of moral judgments on diagnosis by providing more information about the disease: one might have thought that it is only when we have little information about a condition that incidental factors can influence disease judgments, as ambiguity is said to leave room for the influence of social biases on many forms of judgments. As it turns out, however, we failed to find evidence for this intuitive idea. Thus, engineering the deployment of the concept of disease cannot appeal to this intuitive proposal. Furthermore, any proposal about how to remedy the influence of moral judgments on diagnosis (or, to generalize, any other influence on any other concept) should be examined empirically.

### **6. Limitations**

### Our research, while providing critical insights into the effects of perceived moral character on health-related judgments, has several limitations that warrant caution when interpreting the findings. First, the conditions selected for the study, while covering both physical and mental health issues, do not encompass all possible health conditions that might be affected by moral judgments. Exploring a broader array of conditions could help clarify the boundaries of the pathologization and depathologization effects. Second, while factors such as the age and gender of the participants were controlled, we did not investigate cultural variation, which could play a significant role in shaping how moral character impacts health-related judgments and influence the applicability of the results to more diverse populations. Cross-cultural studies are essential to determine if these effects are consistent across different societal norms and values. Third, our study focused solely on moral evaluations elicited by an agent's character. It remains uncertain whether our findings would generalize to evaluations based on other attributes, such as racial identity or socioeconomic class. Fourth, the study utilized brief vignettes in an experimental design, making it unlikely that this setup would provoke the kind of emotionally charged moral evaluations encountered in real-life situations. Therefore, the moderate effect observed should not be interpreted as indicative of the full potential impact that moral evaluations might have outside the laboratory. In more realistic, emotionally intensive contexts, their influence on health and disease judgments could be far more substantial. Lastly, our study only included lay participants, leaving it unresolved whether these findings would extend to health professionals.

**Conclusion**

Building on prior research that identified a pathologization effect, this study further investigated the impacts of perceived moral character on health-related judgments. The study confirmed prior findings of a pathologization effect and revealed that the effects extend deeper, with moral evaluations influencing both person-level and subsystem-level health-related judgments. This suggests a pervasive impact of moral perceptions on health judgments, indicating that moral judgments have an effect across different levels of assessment.

Contrary to our predictions, our findings did not support the depathologization hypothesis. Interestingly, saint-like characters were not only comparable to ordinary characters in health judgments but were also indistinguishably judged from morally evil characters in disease assessments. This unexpected pattern suggests a selective effect of just-world beliefs, potentially influenced by a negativity bias and the processing of incongruent information.

Our findings highlight the need for further research to understand and mitigate the influence of perceived moral character on health judgments, particularly regarding their potential to impact the fairness and objectivity of clinical decisions and patient care. One notable result in this regard is that providing detailed information about the causal mechanisms beyond mere symptoms does not significantly alter judgments. However, with this result, as with the others, it is crucial to investigate to what extent they hold true in real-world interactions and to explore potential mitigation strategies.

**References**

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*, 556–574.

Alicke, M. D., Davis, T. L., & Pezzo, M. V. (1994). A posteriori adjustment of a priori decision criteria. *Social Cognition, 12*, 281–308.

Anderson, V. N. (1992). For whom is this world just?: Sexual orientation and AIDS. *Journal of Applied Social Psychology*, *22*(3), 248-259.

Barnes, E. (2023). *Health Problems: Philosophical Puzzles about the Nature of Health*. Oxford University Press.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, *5*(4), 323-370.

Bizer, G. Y., Hart, J., & Jekogian, A. M. (2012). Belief in a just world and social dominance orientation: Evidence for a mediational pathway predicting negative attitudes and discrimination against individuals with mental illness. *Personality and individual differences*, *52*(3), 428-432.

Braman, A. C., & Lambert, A. J. (2001). Punishing individuals for their infirmities: Effects of personal responsibility, just-world beliefs, and in-group/out-group status. *Journal of Applied Social Psychology*, 31, 1096–1109

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford University Press.

Fischer, E. (2020). Conceptual control: On the feasibility of conceptual engineering. *Inquiry*, 1-29.

Furnham, A. (2003). Belief in a just world: Research progress over the past decade. *Personality and Individual Differences*, *34*(5), 795-817.

Gagné-Julien, A. M. (2024). Beyond conceptual analysis: Social objectivity and conceptual engineering to define disease. In *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* (Vol. 49, No. 2, pp. 147-159).

Guglielmo, S., & Malle, B. F. (2019). Asymmetric morality: Blame is more differentiated and more extreme than praise. *PloS one*, *14*(3), e0213544.

Haslanger, S. (2000). Race and gender:(What) are they?(what) do we want them to be. *Noûs*, *34*(1), 31-55.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, *106*(11), 587-612.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190-194.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and brain sciences*, *33*(4), 315-329.

Kõiv, R. (2024). To reform or to eliminate an attractor?. *Synthese*, *204*(2), 52.

Lerner, M. J. (1977). The justice motive: some hypotheses as to its origins and forms. *J. Pers.* 45, 1–52.

Lerner, M. J. (1980). "The belief in a just world," in *The Belief in a Just World: A Fundamental Delusion*, ed. M. J. Lerner (New York, NY: Springer), 9–30.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological bulletin*, *85*(5), 1030.

Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.

Montada, L. (1998). Belief in a just world: A hybrid of justice motive and self-interest. In L. Montada & M. J. Lerner (Eds.), *Responses to victimizations and belief in a just world.* New York: Plenum Press.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European review of social psychology*, *1*(1), 33-60.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, *5*(4), 296-320.

Schwartz, P. H. (2007). Decision and discovery in defining ‘disease’. *Establishing medical reality: Essays in the metaphysics and epistemology of biomedical science*, 47-63.

Shoemaker, D. (2024). *The Architecture of Blame and Praise: An Interdisciplinary Investigation*. Oxford University Press.

Siegel J. Z., Crockett M. J., & Dolan R. J. (2017). Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition*, 167, 201–211.

Varga, S., & Latham, A. J. (2024a). Is health the absence of disease?. *Inquiry*, 1-21.

Varga, S., & Latham, A. J. (2024b). What is mental health and disorder? Philosophical implications from lay judgments. *Synthese*, *203*(5), 162.

Varga, S. Latham, A.J. & Machery, E. (ms.) The Wicked and the Ill.

1. The term “ordinary” moral character refers to individuals whose choices and actions generally meet moral expectations, without significantly surpassing or falling short of them. Such individuals are neither exceptionally good nor bad and are neither particularly admired nor despised. [↑](#footnote-ref-1)
2. While this hypothesized effect is here outlined in general terms, it is to be expected that its manifestation will also depend on the extent to which the evaluated condition affects responsibility-related capacities. [↑](#footnote-ref-2)
3. MANOVA is a statistical procedure for comparing multivariate sample means. [↑](#footnote-ref-3)
4. While the discussion here has been framed in terms of moral character, the asymmetry might also be explained if people are evaluating matters in terms of an agent’s actions. Thus far we have said that the pathologization effect is being driven by the fact that people think that generalized anxiety, for instance, is a deserving punishment for being a bad person. However, people might instead (or perhaps in addition) think that such a condition is just the *direct* result of that person’s bad actions. People, however, arguably, may not see the connection between morally permissible ordinary or good actions and resulting health conditions in the same light. Exploring the distinction between judgments about actions and character would be a good avenue for future research. [↑](#footnote-ref-4)