UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Mapping the Folk Concept of Health

Permalink

https://escholarship.org/uc/item/90x7x71g

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Huber, Lukas S.
Varga, Somogy
Reuter, Kevin

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Mapping the Folk Concept of Health

Lukas S. Huber[1], **Kevin Reuter**[2] **& Somogy Varga**[3]
[1]University of Bern & University of Tübingen
[2]University of Gothenburg
[3]Aarhus University

Abstract

Health is widely treated as multidimensional, yet little is known about how these dimensions are structured in lay thinking or how this structure guides health-related judgments. We used a conceptual scaling approach to derive *participant-specific* conceptual maps positioning the term *unhealthy* relative to three clusters reflecting Disease, Lifestyle, and Functional Ability aspects of health. Participants' conceptual understanding of *unhealthy* was most closely aligned with a Lifestyle interpretation of health. We also observed substantial inter-individual differences in the degree to which participants' understanding of health was pluralistic. Alignment in participants' conceptual maps predicted how they applied the concept in a subsequent vignette task, suggesting that the structure of lay concepts constrains concept application. These findings may inform psychological theories of health, philosophical debates about its nature, and have implications for effective health communication.

Keywords: health; conceptual scaling; concepts; lifestyle, disease.

Introduction

Few concepts play as central a role in ethical, medical, and political reasoning as *health*, and yet its ordinary meaning remains elusive. Clarifying how ordinary people conceptualize *health* and *unhealthiness* (and whether these conceptions diverge from professional ones) is essential for both empirical research and philosophical theorizing, as such differences shape how patients describe symptoms, how clinicians interpret them, and how policymakers design interventions.

Researchers across philosophy, medicine, and public health have long been interested in what health is, but their discussions have traditionally been polarized. One line of debate concerns whether health should be understood negatively, as the *absence of disease or dysfunction*, or positively, as the *presence of well-being or flourishing*. Another divides naturalist or descriptive accounts from normative and holistic views, which link it to *well-being* or *goal fulfillment* (for reviews, see Kingma, 2019; Murphy, 2021). Yet both lines of debate typically assume that "health" designates a single, unified property.

Empirical work in sociology and health psychology has shown that ordinary conceptions of health are indeed multidimensional. Lay reasoning typically integrates several coexisting aspects of health, as documented across qualitative, survey-based, and corpus-based studies (e.g., Baumgartner et al., n.d.; Bishop & Yardley, 2010; Blaxter, 1990; Downey & Chang, 2013; Herzlich, 1973; Hughner & Kleine, 2004; Millstein & Irwin, 1987; Williams, 1983). For instance, in a large population survey, Blaxter (1990, 2010) identified nine distinct ways people describe being healthy, which do not necessarily exclude each other: absence of disease, health despite disease, a reserve, physical fitness, healthy lifestyle, vitality, good social relationships, ability to do things, and psychosocial well-being (for similar conclusions, see Bishop & Yardley, 2010; Hughner & Kleine, 2004). While such qualitative work tells us what people *explicitly say* about health, recent experimental work sheds light on which features are essential to the concept or how these features trade off when they conflict. For example, Varga and Latham (2024) as well as Reuter et al. (2025) show that lifestyle behaviors like diet, exercise, and weight often outweigh biomedical cues and even disease in health judgments, confirming a distinct lifestyle aspect in lay concepts of health.

Together, these studies suggest that the folk concept of health is organized around at least three separable but interacting ways of thinking about health: a **Disease aspect**, concerned with the presence or absence of pathology and dysfunction; a **Lifestyle aspect**, defined by the presence or absence of health-promoting behaviors and self-care; and a **Functional-ability aspect**, reflecting the presence or absence of injury and the capacity to act or work. But while these findings show how people apply the concept of health, they leave unclear whether those aspects can be identified in people's conceptual representations, how strongly they are associated, and how consistently individuals organize them. Clarifying this structure matters because it determines whether the folk concept of *health* is unified or pluralistic, a question that bears directly on psychological and philosophical theories of health as well as on communication between professionals and the public.

To address this, the present study complements vignette-based designs by employing a conceptual scaling approach (Dranseika, 2024; Huber et al., 2026) that shifts focus from how health concepts are applied to how they are cognitively organized. In particular, we investigate the semanitc relations among the concept "unhealthy" and three clusters of related concepts, each reflecting one of the main ways of thinking about health: *Disease*, *Lifestyle*, and *Functional Ability*. Participants complete a triplet similarity task in which they judge which of two terms is more similar in meaning to a third. These judgments are used to derive participant-specific con-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

ceptual maps in which spatial distance reflects perceived semantic similarity, yielding a fine-grained and largely context-independent characterization of how unhealthy is conceptually organized relative to the three clusters. We then examine whether these individualized conceptual maps predict health-related judgments in a contextualized vignette task.

This approach allows us to investigate the following research questions and corresponding hypotheses.

1. **How does participants' understanding of "unhealthy" align with the three focal aspects of health?** Based on the previous research described above, we hypothesize significant differences in alignment (H1a). In particular, we expect the alignment to be highest for the Lifestyle compared to Disease and Functional Ability (H1b). Both H1a and b were preregistered via the Open Science Framework (OSF).
2. **How pluralistic is participants' understanding of "unhealthy" irrespective of their primary alignment?** Are there large inter-individual differences in how pluralistic participants understand "unhealthy"? We treat these as exploratory research questions and thus did not preregister any corresponding hypotheses.
3. **Is participants' conceptual understanding of "unhealthy"** (as measured by conceptual scaling) **predictive of how people use this concept in a more contextual scenario** (the vignette task)**?** Assuming that conceptual understanding governs aspects of usage, we expect the predictive power of the conceptual maps to be significantly better than random prediction of vignette responses (H2; preregistered on OSF).

Before we describe the conducted study to address these research questions and evaluate the corresponding hypothesis, we proceed by briefly describing conceptual scaling—the central paradigm applied in the present work.

Conceptual Scaling

How can we investigate the position of *health* in conceptual space with respect to the three focal health aspects: Disease, Lifestyle, and Functional Ability? Traditionally, researchers have relied on vignette-based paradigms to elicit judgments about people's conceptions of health. How can we investigate the position of *health* in conceptual space with respect to the three focal health aspects: Disease, Lifestyle, and Functional Ability? Traditionally, researchers have relied on vignette-based paradigms to elicit judgments about people's conceptions of health. However, such methods are ill-suited for this task: vignette studies typically probe categorical or evaluative responses to isolated cases, thereby presupposing specific theoretical framings and obscuring the global similarity structure of the concept. As a result, they offer limited insight into how *health* is systematically related to alternative conceptions across a broader conceptual space.

We therefore turn to a method with its roots in psychophysics. Scaling techniques have long been used to uncover the structure of mental representations from behavioral data encoding perceived similarity relations (for a recent review, see Roads & Love, 2024). In a canonical paradigm, participants provide similarity judgments among stimuli, which are then used to recover a spatial configuration in which shorter distances between items correspond to greater perceived similarity (e.g., Gescheider, 2013a, 2013b; Shepard, 1957, 1962; Torgerson, 1952). Beyond perceptual spaces, scaling methods have also been extended to the investigation of concepts, relying on linguistic labels rather than psychophysical stimuli to elicit similarity judgments (e.g., Arnold, 1971; Henley, 1969; Marti et al., 2023; Waraich & Victor, 2023). We refer to these approaches as **conceptual scaling** (as opposed to perceptual scaling) and to the resulting spatial embeddings as **conceptual maps**.

Although most applications of *conceptual* scaling have focused on comparatively concrete concepts (e.g., animals, fruits, tools), a small number of studies (e.g., Brandt, 2022; Dranseika, 2024) have started to examine genuinely abstract ones. Recent work on *truth* by Huber et al. (2026) shows that conceptual scaling can uncover structural features of abstract concepts that challenge standard philosophical accounts: rather than aligning neatly with correspondence or coherence, folk conceptions of truth are shaped by an interconnected cluster of notions, including honesty and authenticity (see also Ricciardi & Reuter, 2025; Zyglewicz et al., n.d.). These findings underscore the promise of conceptual scaling for mapping abstract conceptual structure beyond the reach of traditional philosophical analysis.

Mapping lay people's understanding of health

In this work, we closely follow the approach presented by Huber et al. (2026) and report an online experiment in which participants first rated the similarity between the term "unhealthy" and six linguistic labels (each representing one of three interpretations), and then completed a vignette task. Using the similarity judgments to construct conceptual maps and the responses from the vignette task, we investigate the research questions posed in the Introduction. In particular, we identified each participant's dominant alignment by identifying which of the three different aspects lay closest to unhealthy in their individualized conceptual map(addressing Question 1). We then use this dominant alignment to predict participants' responses in the vignette task (addressing Question 3). Additionally, we calculate a pluralism score to quantify whether unhealthy maintains a balanced spatial relationship to multiple aspects rather than clustering narrowly around a single one (addressing Question 2).

Methods

Triplet task. Stimuli for the triplet task consisted of seven linguistic labels. Besides "unhealthy" we used six other terms, two for each focal health aspect: "ill" & "sick" for Disease, "unfit" & "overweight" for Lifestyle, and "injured" and "impaired" for Functional Ability. These stimuli were selected for three reasons. First, prior work has identified them as reliably

representing the three aspects of health (Blaxter, 2010; Reuter et al., 2025; Varga & Latham, 2024). Second, all of these terms are in common use and occur at roughly comparable frequencies (see, for example, their distribution in the Corpus of Contemporary American English [COCA]). Third, they are non-composite words that can be readily understood without requiring extended reading time.[1] On each trial, participants viewed three terms arranged in a triangular configuration, with one term (the anchor) at the top and two terms (the targets) at the bottom. Participants were prompted to make simple ordinal decisions about which of the two targets is more similar to the anchor. Responses were required within a six-second window using the `left` and `right` arrow keys. The task comprised 97 trials (79 unique triplets plus two repetitions of 9), drawn from the full set of 105 possible combinations ($3\binom{7}{3}$) in line with other studies (e.g., Haghiri et al., 2020; Huber et al., 2026; Künstle et al., 2022).

Vignette task. In the vignette task, we presented participants with the following vignette as introduced by Reuter, Latham & Varga (2025);

> *During the last six months, Erica has not had any diseases like infections or bodily dysfunctions, nor injuries such as fractures.*
> *During the last six months, Erica has also not experienced any symptoms like headaches, fever, or nausea.*
> *During the last six months, Erica has not exercised, slept little, ate poorly, and smoked a lot.*
> *Please tell us how much you agree with the following statement:*
> **During the last six months, Erica was unhealthy.**

Participants indicated their level of agreement using a continuous visual slider scale ranging from –50 (totally disagree) to +50 (totally agree), with zero as a neutral point. For a randomly selected half of the participants, the statement was reverse-coded, with the target word "unhealthy" replaced by "healthy". This allowed us to counterbalance potential response biases related to the wording of the focal statement.

Procedure. From Prolific, participants were forwarded to a survey on Qualtrics to collect demographic information. Participants were then directed to an online PsychoPy experiment (Peirce et al., 2019) hosted on Pavlovia. Prior to the main phase of the triplet task, they completed two short practice sessions (10 trials each) to familiarize themselves with the triplet procedure. The 97 experimental trials were presented in blocks of 19. Once they completed all triplet trials, they were directed back to Qualtrics, where they read and responded to the vignette as outlined above.

[1]This is a main reason why we focus on "unhealthy" and not it's positive counterpart "healthy": positive terms for the absence of disease, a healthy lifestyle, or intact functional ability tend to be often technical, or expressed as composite constructions rather than simple lexical items (e.g., *asymptomatic*, *fully functional*, *physically robust*).

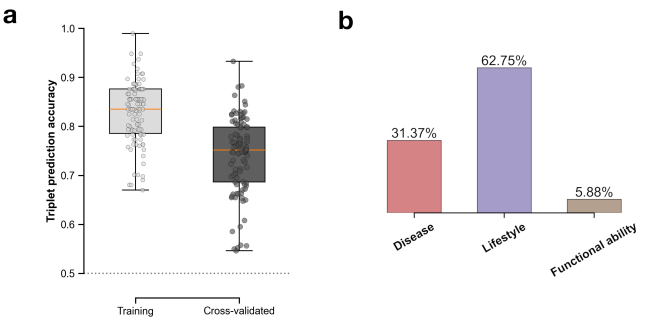

Figure 1: **Participants' conceptual maps are most aligned with a Lifestyle interpretation off health.** (**a**) Training and cross-validated triplet accuracies demonstrate that the 2D conceptual maps provide a robust fit to participants' triplet judgments. Boxplots display the interquartile range; horizontal orange lines indicate the median. (**b**) When assigning each participant to the cluster closest to unhealthy, most conceptual maps align most strongly with Lifestyle, with significantly fewer aligned with Disease and only a very small number with Functional Ability.

Participants. We recruited 160 participants (M_{age} = 41.95, SD_{age} = 13.55; gender distribution (81 female / 79 male / 0 other)) from the UK and the United States via the online platform Prolific. We obtained informed consent from all participants. Due to issues in our data-matching procedure, we were only able to match the triplet data of 130 participants to the corresponding vignette data.[2] As detailed in the preregistration, participants were excluded if they rushed through the triplet task in less than five minutes (9 participants) and if the non-target terms did not cluster as expected (19 participants; see Supplementary Materials). The final sample of 102 participants (M_{age} = 42.68, SD_{age} = 13.44; gender distribution (49 female / 53 male / 0 other)) did not statistically differ from the initial sample regarding age (Welch's $t = \breve{}0.44, p = 0.629$), gender ($\chi^2(1) = 0.079, p = 0.779$) and education ($\chi^2(4) = 0.433, p = 0.980$).

Analyses & Results

Individual conceptual maps reliably capture participants' similarity judgments. Following Huber et al. (2026), we used the soft ordinal embedding algorithm (Terada & Luxburg, 2014) (as implemented in the Python toolbox `cblearn`) to embed the stimuli into a 2D conceptual map. Within those maps, the distances correspond inversely to the perceived similarity among stimuli (terms that are perceived as

[2]The mismatch resulted from a procedural weakness in our study setup overlooked while piloting: participants were asked to *manually* transfer their Prolific and Pavlovia codes between platforms rather than having this step automated. As a consequence, some participants entered incorrect codes, and their data could not be matched across the triplet and vignette tasks.

similar should be represented close together, and vice versa). Notably, the number of triplet judgments collected per participant is sufficient to obtain a conceptual map for each participant.

To quantify how well a given conceptual map captured the triplet responses of the corresponding participant, we first computed triplet accuracy, defined as the proportion of triplets for which the distances in the recovered conceptual map agree with the participant's judgments.[3] Across participants, mean triplet accuracy was $M = .83$ ($SD = .06$), indicating that the 2D embeddings captured a substantial portion of each participant's similarity structure (chance would be 0.5). As a more conservative measure of goodness-of-fit, we also computed cross-validated triplet accuracy using a 5×5 repeated 5-fold procedure, in which the embedding was repeatedly estimated from a training split and evaluated on held-out triplets. Mean cross-validated accuracy was $M = .74$ ($SD = .08$), reflecting robust generalization of the embeddings beyond the specific triplets used to construct them. A summary of both training and cross-validated accuracies is shown in Figure 1.

People most often understand health in terms of Lifestyle. To identify which health aspect each participant's conceptual map aligns with most closely (and thus to evaluate hypotheses H1a & H1b; see Section), we computed the centroid of each dimension based on its defining terms (e.g., "unfit" and "overweight" for Lifestyle), treating this centroid as the theory's estimated position within the conceptual map. We then assigned each participant to the aspect whose centroid was nearest to "unhealthy". The resulting "winner-takes-all" distribution is shown in Figure 1. Alignment was strongest for Lifestyle (62.75%; $n = 64$), followed by Disease (31.37%, $n = 32$) and Functional Ability (5.88%; $n = 6$).

A χ^2 goodness-of-fit test against a uniform $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ distribution confirmed that these proportions deviate strongly from chance ($\chi^2(2) = 55.09$, $p < .0001$), in line with H1a. Follow-up one-sided binomial tests—each Bonferroni-corrected to $\alpha = 0.05/3 \approx 0.017$—revealed that "unhealthy" was significantly more often closest to Lifestyle than to Disease ($p = .0007$), significantly more often closest to Lifestyle than to Functional Ability ($p < .0001$), and significantly more often closest to Disease than to Functional Ability ($p < .0001$). These results provide clear support for H1b, which predicted that alignment with the Lifestyle cluster would occur more frequently than with the second-highest cluster.

Lay concepts of health range from monistic to lifestyle–disease pluralism. To explore how pluralistically participants understand health, we computed a pluralism score for each participant. Intuitively, this score reflects how evenly unhealthy is positioned relative to the three health aspects in the participant's conceptual map. A participant scores **low** on pluralism if unhealthy lies very close to one dimension (e.g., Disease: *ill/sick*) and far from the others, indicating a monistic understanding. In contrast, a participant scores **high** on pluralism if unhealthy is positioned at comparable distances to two or all three clusters, indicating that multiple aspects contribute similarly to their understanding of health. See Supplementary Material for details on how we compute pluralism scores.

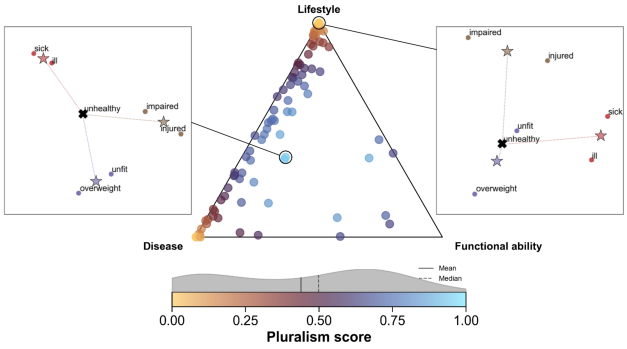

Figure 2: **Conceptual maps reveal heterogeneous degrees of pluralism in lay concepts of health.** The barycentric in the center plot summarizes where participants position unhealthy relative to the three focal health aspects. Points near one corner indicate monistic understandings; points between corners indicate pluralism. Insets show two representative individual conceptual maps illustrating the most common patterns in our data: a strongly monistic Lifestyle-aligned conceptual map (left) and a pluralistic conceptual map blending Lifestyle and Disease (right). Stars in each inset represent the estimated centroids of each aspect, derived from the positions of their two defining terms. The marginal density above the colorbar shows the overall distribution of pluralism scores.

Figure 2 visualizes participants' pluralism scores by locating each participant's conceptual map in the simplex defined by the three health aspects. Each point represents a participant, positioned according to the relative proximity of unhealthy to the three aspects. Participants whose maps place unhealthy very close to a single cluster appear near that cluster's "corner", whereas participants with more balanced proximities appear further toward the center.

The pluralism analysis reveals that the above-reported winner-takes-all alignment does not imply that participants hold a unitary understanding of unhealthy. Using the optimal smoothing parameter ($\gamma^* = 5.59$; see Supplementary Material), pluralism values ranged widely, with a mean of 0.45 and a median of 0.51. While some individuals align strongly with one aspect—visible as points near a corner in Figure 2, indicating a monistic understanding—many participants are positioned between Lifestyle and Disease, reflecting meaningful contributions from both aspects.

Conceptual maps predict how people apply health in context. To test whether conceptual understanding predicts how

[3] Triplet accuracy is the proportion of trials for which this embedding distances predict the participant's actual triplet response.

people apply the concept, we first recoded the inverse polarity of the vignette responses from the participants in the "healthy-framing" condition. Then we transformed the responses into a binary format, with values < 0 on the slider scale coded as 0 (no, Erica was not unhealthy; aligning with Disease/Functional Ability) and values > 0 coded as 1 (yes, Erica was unhealthy; aligning with Lifestyle).

Next, we used participants' dominant alignment to derive binary predictions. If a participant is aligned with Lifestyle, we predict that they judged Erica to be unhealthy; otherwise, if aligned with Disease or Functional Ability, we predict that they judged Erica not to be unhealthy. Across all participants, we thus predicted that 62.75% ($n = 64$) would report that Erica was unhealthy and 37.25% ($n = 38$) that Erica was healthy.

We find that 76.47% ($n = 78$) of participants indicated that Erica was unhealthy, whereas 23.53% ($n = 24$) reported that Erica was not unhealthy. The conceptual maps predict a majority "unhealthy" responses, but underestimate the magnitude. Given this mismatch in predicted versus observed base rates, the maximal achievable individual-level prediction accuracy was 86.27%. Despite this constraint, the conceptual maps correctly predicted 60.80% ($n = 62$) of participants' responses (this corresponds to ~70.5% of the maximum achievable accuracy given the base-rate mismatch). A one-sided binomial test against chance (50%) confirmed above chance performance ($p = 0.0186$), supporting H2. Importantly, these predictions were generated without fitting any model to the vignette data. Instead, we relied solely on the geometry of each participant's conceptual map derived from similarity judgments to produce zero-shot binary predictions of *individual* responses.

In a non-preregistered exploratory analysis, we examined whether prediction accuracy improved for participants who held stronger beliefs about Erica's health. Since responses in the vignette task were provided on a visual slider ranging from –50 (totally disagree) to +50 (totally agree), we used the absolute value of each participant's rating as a proxy for belief strength. Participants were then ranked according to this measure, and progressively larger subsets were formed in 10% increments—once by adding participants with higher belief strength first, and once by adding participants with lower belief strength first (see Figure 3). For each subset, we computed prediction accuracy using a bootstrap procedure, resampling participants with replacement to the full sample size and estimating mean accuracy along with a 95% confidence interval. This step-wise inclusion procedure allowed us to test whether belief strength is systematically related to predictability while holding subset size constant. The results showed that prediction accuracy was highest for participants with stronger beliefs, peaking at over 70%, whereas accuracy for participants with the weakest beliefs did not exceed chance level (compare subplots in Figure 3).

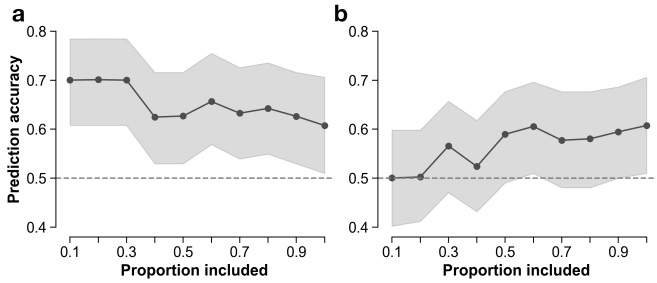

Figure 3: **Prediction accuracy varies systematically with belief-strength.** Both traces show bootstrapped prediction accuracy (2,000 bootstraps, solid lines ±95% CI, shaded) as progressively larger subsets of participants are included. (**a**) Participants are ordered from strongest to weakest belief strength, so the 0.1 point reflects the 10% with the strongest beliefs. Prediction accuracy is highest if beliefs are strong and gradually declines as weaker-belief participants are added. (**b**) Participants are ordered from weakest to strongest belief strength, with the 0.1 point reflecting the 10% with the weakest beliefs. Prediction accuracy begins at chance level and increases as participants with stronger beliefs are included. The dashed horizontal line indicates chance prediction accuracy (0.50).

Discussion

Earlier work has argued that health is a multidimensional concept with at least three important aspects: Disease, Lifestyle, and Functional Ability. In the present study, we used conceptual scaling to examine how the lay concept *unhealthy* is organized relative to these three aspects, and to test whether this underlying conceptual structure predicts how people apply the concept. Supporting our first two hypotheses, the distribution of alignments differed markedly from chance (H1a), and the Lifestyle emerged as the dominant point of reference (H1b): most participants' maps placed *unhealthy* closest to Lifestyle (62.75%), with fewer aligning with Disease (31.37%) and very few with Functional Ability (5.88%).

While some participants tended to a monistic understanding of health, we also found clear evidence of pluralism, as participants drew on multiple aspects when thinking about the concept. Pluralistic conceptual maps most often combined Lifestyle and Disease, while Functional Ability played only a very minor role. Consistent with this pattern, no participant showed a purely functional-ability-based understanding. One way to interpret this pattern is through the framework of dual character concepts, where a concept incorporates both a descriptive and a normative dimension, (e.g., Knobe et al., 2013; Reuter, 2019). On such an interpretation, ordinary judgments about health may simultaneously track biomedical status (e.g., disease or dysfunction) and more evaluative considerations tied to lifestyle. The observed combination of Lifestyle and Disease aspects in many participants' conceptual maps may therefore reflect not merely conceptual polysemy, but the coexistence of partially independent dimensions within the folk concept of health itself.

Quantifying the structural relations among different health aspects in individuals' conceptual maps enabled us to test whether this structure is stable and consequential. Supporting H2, participant-specific conceptual maps predicted vignette judgments above chance. A zero-shot mapping from each participant's conceptual alignment to their vignette response correctly predicted 60.80% of individual judgments, with accuracy exceeding 70% among participants who expressed stronger evaluations about Erica's health status. Notably, these predictions required no fitting to vignette data and relied solely on the geometry of each participant's conceptual map. This finding indicates that conceptual structure meaningfully constrains how individuals apply the concept a novel judgment context.

Our findings advance existing qualitative and experimental work on the concept of health in two ways. First, complementing prior qualitative studies show that lay understandings of health are multidimensional, encompassing disease status, functional ability, and lifestyle (e.g., Bishop & Yardley, 2010; Blaxter, 1990; Herzlich, 1973; Hughner & Kleine, 2004). Our results refine this picture by quantifying the conceptual structure of *unhealthy* at the level of individual participants, resulting in participant-specific conceptual maps rather than relying on verbally articulated structures. We show that the multidimensionality often attributed to folk concepts of health is not evenly expressed: Lifestyle and Disease aspects predominate.

Second, extending prior experimental work, we provide *context-independent* evidence about the anchoring of everyday *unhealthy* judgments. Conceptual maps systematically located *unhealthy* closest to Lifestyle terms, rather than Disease or Functional Ability, indicating that lifestyle considerations form the primary conceptual anchor for negative health judgments. This pattern strengthens prior findings suggesting that people weigh behavioral factors more strongly than disease or particular biomedical cues (Reuter et al., 2025; Varga & Latham, 2024).

In addition to advancing qualitative and experimental work, our findings have potential implications both for philosophical debates about the nature of health and for clinical and public health communication. For philosophical debates, our asymmetrical pattern challenges the influential disease-based view, which holds that health is nothing more than the absence of disease. Although we do not take this to settle how health *ought* to be defined, theories taking ordinary language as a constraint must accommodate this broader structure. For health communication, if laypeople tend to organize *unhealthy* primarily around lifestyle, while clinicians frame health status mainly in terms of pathology, there is a plausible risk of misalignment in how information about risk, diagnosis, and prognosis is interpreted. Our study does not directly test such possible effects, but it points to structural features of the *unhealthy* concept that practitioners should be mindful of when designing communicative strategies.

Limitations. Several limitations related to our stimulus choice constrain our conclusions and point to future work. First, our operationalization of the "Functional Ability/Damage" aspect is relatively narrow. The terms *injured* and *impaired* likely emphasize acute bodily harm more than everyday or long-term functioning, which may help explain the weak role this dimension played in participants' maps. A broader stimulus set including mental, social, and everyday functional capacities would provide a stronger test. Second, all six comparison terms primarily apply to living beings, whereas *unhealthy* is also commonly used for non-living entities such as food, lifestyles, or environments. This mismatch may have biased interpretations toward biological or bodily cases. Third, some cluster terms differ in semantic granularity: *sick* and *ill* are near-synonyms, whereas *unfit* and *overweight* denote more specific forms of unhealthiness. Although this asymmetry reflects ordinary health language, future work should examine whether alternative term selections produce different clustering or weighting patterns.

A second set of limitations concerns the broader scope of this study. We focused specifically on the negative term *unhealthy*, rather than on the positive term *healthy* or other negative locutions such as *not healthy*. It is possible that these expressions are applied to systematically different cases, and our results should therefore not be assumed to generalize across all linguistic expressions of health status. Moreover, although social and mental dimensions are widely regarded as central to health, our stimulus set was designed primarily to capture physical aspects, in line with dominant theoretical debates. This focus leaves open how strongly psychosocial considerations structure the concept of *unhealthy* in more ecologically rich contexts. In addition, our conceptual-scaling task relied on a deliberately small set of seven terms closely tied to theory. Future work should test whether the observed weighting patterns remain stable with richer stimulus sets and additional dimensions of health. Moreover, Since our sample included only English-speaking participants, cross-cultural and cross-linguistic studies will be needed to assess the generality of the asymmetrically weighted structure observed here.

Conclusion

This study used conceptual scaling to illuminate the structure of ordinary health concepts with a level of precision that complements and extends existing qualitative and experimental work. Laypeople draw on multiple aspects when thinking about *unhealthy*, but these aspects are weighted asymmetrically: lifestyle considerations dominate, disease plays a secondary role, and functional ability contributes comparatively little. Crucially, participant-specific conceptual maps predicted how individuals applied *unhealthy* in a concrete scenario without any task-specific fitting, indicating that conceptual organization itself constrains concept use. These findings reveal a pluralistic but asymmetrical structure, with implications both for philosophical theorizing about health and for practical questions of health communication.

Code and Data Availability. All empirical data from both the triplet and vignette tasks, as well as the code used to analyze these data and generate the figures reported in this work, are available at: https://github.com/lukasShuber/map_health.git
Supplementary Material. https://github.com/lukasShuber/map_health/blob/main/supplementary_materials.pdf

Acknowledgments

During the preparation of this work the authors used ChatGPT (GPT version 5.1 and 5.2) in order to brainstorm for alternative wordings for some specific sentences. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. Lukas S. Huber was funded by the Swiss National Science Foundation (grant number: P000PS_214659).

References

Arnold, J. B. (1971). A multidimensional scaling study of semantic distance. *Journal of Experimental Psychology*, *90*(2), 349.

Baumgartner, L., Reuter, K., & Varga, S. (n.d.). *A common language? how laypeople and medical professionals use health-related vocabulary in online advice* [Manuscript in preparation].

Bishop, F. L., & Yardley, L. (2010). The development and initial validation of a new measure of lay definitions of health: The wellness beliefs scale. *Psychology and Health*, *25*(3), 271–287.

Blaxter, M. (1990). *Health and lifestyles*. Routledge.

Blaxter, M. (2010). *Health*. Polity.

Brandt, M. J. (2022). Measuring the belief system of a person. *Journal of Personality and Social Psychology*, *123*(4), 830.

Downey, C. A., & Chang, E. C. (2013). Assessment of everyday beliefs about health: The lay concepts of health inventory, college student version. *Psychology and Health*, *28*(7), 818–832.

Dranseika, V. (2024). Folk beliefs about phenomenological differences and similarities between kinds of mental states. In D. Gregory & K. Michaelian (Eds.), *Dreaming and memory: Philosophical issues* (pp. 325–352). Springer International Publishing, Cham.

Gescheider, G. A. (2013a). The Measurement of Sensory Attributes and Discrimination Scales. In *Psychophysics: The Fundamentals* (pp. 183–206). Taylor & Francis.

Gescheider, G. A. (2013b). Partition Scales. In *Psychophysics: The Fundamentals* (pp. 183–206). Taylor & Francis.

Haghiri, S., Wichmann, F. A., & von Luxburg, U. (2020). Estimation of perceptual scales using ordinal embedding. *Journal of Vision*, *20*(9), 14. https://doi.org/10.1167/jov.20.9.14

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 176–184. https://doi.org/10.1016/S0022-5371(69)80058-7

Herzlich, C. (1973). *Health and illness*. Academic Press.

Huber, L. S., Künstle, D.-E., & Reuter, K. (2026). Tracing truth through conceptual scaling. *Cognition*, *266*, 106321.

Hughner, R. S., & Kleine, S. S. (2004). Views of health in the lay sector: A compilation and review of how individuals think about health. *Health*, *8*(4), 395–422.

Kingma, E. (2019). Contemporary accounts of health. In E. Kingma (Ed.), *Health* (pp. 289–318). Oxford University Press.

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, *127*(2), 242–257.

Künstle, D.-E., von Luxburg, U., & Wichmann, F. A. (2022). Estimating the perceived dimension of psychophysical stimuli using triplet accuracy and hypothesis testing. *Journal of Vision*, *22*(13), 5–5.

Marti, L., Wu, S., Piantadosi, S. T., & Kidd, C. (2023). Latent diversity in human concepts. *Open Mind*, *7*, 79–92.

Millstein, S. G., & Irwin, C. E. (1987). Concepts of health and illness: Different constructs or variations on a theme? *Health Psychology*, *6*(6), 515–.

Murphy, D. (2021). Concepts of disease and health. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2021). https://plato.stanford.edu/archives/spr2021/entries/health-disease/

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195–203.

Reuter, K. (2019). Dual character concepts. *Philosophy Compass*, *14*(1), e12557. https://doi.org/10.1111/phc3.12557

Reuter, K., Latham, A. J., & Varga, S. (2025). Concept(s) of health: Lifestyle at the heart of modern health. *Erkenntnis*, 1–26.

Ricciardi, G., & Reuter, K. (2025). Truth and perspective. *Philosophical Studies*, 1–26.

Roads, B. D., & Love, B. C. (2024). Modeling similarity and psychological space [PMID: 37562499]. *Annual Review of Psychology*, *75*(1), 215–240. https://doi.org/10.1146/annurev-psych-040323-115131

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. i. *Psychometrika*, *27*(2), 125–140.

Terada, Y., & Luxburg, U. (2014). Local ordinal embedding. *International Conference on Machine Learning*, 847–855.

Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, *17*(4), 401–419.

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

Waraich, S. A., & Victor, J. D. (2023). The geometry of low- and high-level perceptual spaces. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1460-23.2023

Williams, R. (1983). Concepts of health: An analysis of lay logic. *Sociology*, *17*(2), 185–205.
Zyglewicz, T., Reuter, K., & Mandelbaum, E. (n.d.). Truth-value judgments [Forthcoming]. *Philosophical Studies*.