Effect of Assessment Method on the Discrepancy between Judgments of Health Disorders People Have and Do Not Have: A Web Study

Jonathan Baron  
*University of Pennsylvania*

David A. Asch  
*University of Pennsylvania*

Angela Fagerlin

Christopher A. Jepson  
*University of Pennsylvania*

George Loewenstein

*See next page for additional authors*

Follow this and additional works at: https://repository.upenn.edu/marketing_papers

Part of the Marketing Commons, Medical Humanities Commons, and the Statistics and Probability Commons

**Recommended Citation**


This paper is posted at ScholarlyCommons. https://repository.upenn.edu/marketing_papers/272

For more information, please contact repository@pobox.upenn.edu.
Effect of Assessment Method on the Discrepancy between Judgments of Health Disorders People Have and Do Not Have: A Web Study

Abstract
Three experiments on the World Wide Web asked subjects to rate the severity of common health disorders such as acne or arthritis. People who had a disorder ("Haves") tended to rate it as less severe than people who did not have it ("Not-haves"). Two explanations of this Have versus Not-have discrepancy were rejected. By one account, people change their reference point when they rate a disorder that they have. More precise reference points would, on this account, reduce the discrepancy, but, if anything, the discrepancy was larger. By another account, people who do not have the disorder focus on attributes that are most affected by it, and the discrepancy should decrease when people make ratings on several attributes. Again, if anything, the discrepancy increased when ratings were on separate attributes (combined by a weighted average). The discrepancy varied in size and direction across disorders. Subjects also thought that they would be less affected than others.

Keywords
adaptation, utility assessment, response shift, focusing illusion

Disciplines
Business | Marketing | Medical Humanities | Statistics and Probability

Author(s)
Jonathan Baron, David A. Asch, Angela Fagerlin, Christopher A. Jepson, George Loewenstein, Jason Riis, Margaret G. Stineman, and Peter A. Ubel

This technical report is available at ScholarlyCommons: https://repository.upenn.edu/marketing_papers/272
Effect of assessment method on the discrepancy between judgments of health disorders people have and do not have: A Web study.

Jonathan Baron\textsuperscript{1,2} David A. Asch\textsuperscript{2,3,4} Angela Fagerlin\textsuperscript{5} Christopher Jepson\textsuperscript{3} George Loewenstein\textsuperscript{6} Jason Riis\textsuperscript{7} Margaret G. Stineman\textsuperscript{2,8} Peter A. Ubel\textsuperscript{5,9}

July 7, 2002

\textsuperscript{1}Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6196, e-mail baron@psych.upenn.edu, funded by N.I.H. Grant AG16258 (P. Ubel, Principal Investigator)

\textsuperscript{2}Leonard Davis Institute of Health Economics, University of Pennsylvania

\textsuperscript{3}Division of General Internal Medicine, University of Pennsylvania School of Medicine

\textsuperscript{4}Veterans Affairs Medical Center, Philadelphia, Pennsylvania

\textsuperscript{5}Division of General Internal Medicine, University of Michigan

\textsuperscript{6}Department of Social and Decision Sciences, Carnegie-Mellon University

\textsuperscript{7}Department of Psychology, University of Michigan

\textsuperscript{8}Division of Rehabilitation Medicine, University of Pennsylvania School of Medicine

\textsuperscript{9}Veterans Affairs Medical Center, Ann Arbor, Michigan
Abstract

Three experiments on the World Wide Web asked subjects to rate the seriousness of common health disorders such as acne or arthritis. People who had a disorder (“Haves”) tended to rate it as less serious than people who did not have it (“Not-haves”). Two explanations of this Have vs. Not-have discrepancy were rejected. By one account, people change their reference point when they rate a disorder that they have. More precise reference points would, on this account, reduce the discrepancy, but, if anything, the discrepancy was larger. By another account, people who do not have the disorder focus on attributes that are most affected by it, and the discrepancy should decrease when people make ratings on several attributes. Again, if anything, the discrepancy increased when ratings were on separate attributes (combined by a weighted average). The discrepancy varied in size and direction across disorders. Subjects also thought that they would be less affected than others.
1 Introduction

Brickman and his colleagues argued that people adapt to their lives, so that the joy of positive changes and the despair of negative changes both wear off over time [1, 2]. Paraplegics, after a few years, are almost as “happy” as other people. Consistent with these observations, patients with chronic health disorders sometimes do not see their disorder as very bad, compared to judgments offered by those who have not experienced the disorder [3, 4]. In these studies, patients report that their quality of life is significantly better than the public estimates it would be. For example, Sackett and Torrance [4] found that the general public estimates the health related quality of life (HRQoL) of dialysis at a value of 0.39 (on a scale from 0 for conditions as bad as death to 1 for perfect health), whereas dialysis patients estimate their HRQoL at 0.56. Boyd et al. [3] found that patients without colostomies estimate the HRQoL of living with a colostomy at 0.80, while patients with colostomies rate their own HRQoL at 0.92. A similar discrepancy has been seen between rheumatoid arthritis patients and the general public.

Other studies, however, fail to find such discrepancies (e.g., Llewellyn-Thomas et al. [5]; O’Connor et al. [6]). These studies often examine short-term conditions, such as the experience of radiation therapy, evaluated by the same patients, before and during the experience.

The question we address here is whether the discrepancy between the ratings of people who have and do not have a disorder — the Have–Not-have discrepancy — can be found for common disorders in a broad sample of people. We ask this by using questionnaires on the World Wide Web, available to anyone. Although our sample is surely not representative of all human beings (the population of interest), it is diverse, and it includes many people with the health disorders of interest. If the discrepancy can be demonstrated by this sort of method, then future research on it is easier than would otherwise be the case.

We test here two explanations of the Have–Not-have discrepancy. In one, the meaning of the response scale changes when people evaluate disorders that they have. A term such as “good health,” which might be used for the top of a rating scale, can mean one thing to a normal healthy
person — being able to play tennis or ride a bike, for example — and quite another thing to a person who has just lost a leg, where it might mean being able to go back to the office. This can happen either because people compare themselves to others with similar disorders or because they adopt different aspiration levels. In either case, the discrepancy should be reduced by making the response scale more explicit, so that the terms used to name the ends of it are not so subject to variable interpretation.

This explanation is similar to the idea of “response shift” [7]. People undergoing cancer therapy, in some studies, retrospectively evaluate their pre-therapy condition as better than they rated that condition at the time they experienced it [8], or as worse, if their condition improved [9, 10]. Analogous results are found for transplant patients, who rate their pre-transplant quality of life (QOL) as lower after the transplant than they did at the time [11, 12]. Jansen et al. [13], however, found no evidence for a response shift.

Here, in Experiments 1 and 2, we test this explanation by assessing the Have–Not-have discrepancy with different kinds of response scale, which differ in the clarity of their end point. If a scale is defined so that end points have a constant meaning, and the discrepancy is reduced, then the use of undefined scales is part of the explanation of it.

The other explanation holds that the discrepancy is the result of a focusing illusion. When imagining the situation of others who have a disorder, people focus disproportionately on what is affected by the disorder while ignoring those things that are not affected. In the original demonstration of this focusing illusion, students in both California and the Midwest predict that they would be significantly happier living in California than in the Midwest [14]. This prediction correlates strongly with how important they feel weather is to their quality of life. Yet no significant difference in happiness is found between these two groups of students, suggesting that they focus disproportionately on the effect of weather on their quality of life when they compare themselves to those in a different climate.

In an earlier study, Ubel et al. [15] explored whether a focusing illusion contributes to general public estimates of the QOL associated with disabilities. Subjects estimated the QOL of either
paraplegia, below the knee amputation, or partial blindness. Then the experimenters attempted
to “de-focus” subjects by having them reflect on the impact of these disabilities on a wide range
of life domains. This de-focusing task should keep people from thinking too narrowly about the
life domains affected by the disability. For example, subjects were asked to think about how such
a disability would affect their family life, assuming that for most, it would have little effect. This
manipulation showed no effect. The focusing hypothesis was not supported.

Ubel et al. [16] replicated this negative result with three other focusing tasks. One de-focusing
task involved asking people to imagine how the disability in question would affect eight concrete
life events, such as “paying bills and taxes” and “reading or watching TV or movies.” Another
de-focusing task asked people to list those things that took up the largest amount of their time on
the previous day and then indicate how much the disability in question would impact those things.
A third de-focusing task asked people to specifically think not only about things that would be
made worse about the disability in question, but also things that would be unchanged and things
that would be better because of the disability. None of the tasks affected the magnitude of the
discrepancy, measured after doing the task.

Here, we test the focusing illusion in a different way, in Experiment 3, by asking whether the
Have–Not-have discrepancy is reduced when subjects provide ratings of the disorder attribute by
attribute. This method insures that subjects attend to all the attributes we provide. It should thus
prevent focusing, at least to some extent, and reduce the discrepancy, if the focusing explanation
is at work.

Also relevant to the focusing illusion is a comparison of different kinds of scales in Experiments
1 and 2. Scales that concern health should be less subject to a focusing illusion than those that are
broader, such as those concerning quality of life, or happiness. In fact the Brickman study used a
happiness scale. We might expect the happiness scale to show the smallest discrepancy.

In sum, our two main questions are whether the Have–Not-have discrepancy can be reduced by
the use of well-defined scales (Experiments 1 and 2, which also examine different kinds of scales),
and by the use of attribute-by-attribute ratings.
Experiments 1 and 2 also address a subsidiary question: whether people think that they can adapt to a disorder better than others can. In general, people tend to think they are “above average” on all good things [19]. Most studies comparing Have and Not-have are asking the Have group (those with the disorder) to rate themselves and the Not-have group to rate others. Thus, Have–Not-have is confounded with ratings of self and others. The Have–Not-have discrepancy could result from a belief that “I can adapt to this better than other people can,” whether or not the person answering has the disorder or not. We thus ask for ratings of self with the disorder, self without it, others with, and others without. One of the two Self ratings is necessarily hypothetical. A Self-Other difference would support this explanation.

It is also possible that this Self-Other difference is found mainly in those who have the disorder. If so, Self-Other would interact with Have–Not-have.

2 Experiment 1

The main purpose of this experiment was to look for a discrepancy in the ratings of common disorders, in which people who do not have the disorder rate it as more serious than those who have it (Not-have vs. Have). The ratings say how much dis-utility a disorder will create. So our general hypothesis is that Have’s don’t think things are as bad as Not-have’s think they would be. The justification for this hypothesis is that when people have found discrepancies, they have typically been in this direction.

Note that this discrepancy need not always go in this direction. For certain health disorders, especially those that primarily effect mood and subjective well being, we expect the discrepancy to go in the other direction. Classic examples of these types of disorders would be anxiety, depression and pain. And indeed in this study, one of the disorders that involves pain, migraines, showed a discrepancy in the opposite direction (Have rated worse than Not-have). Although we know of no previous findings of a reversed discrepancy of this sort, Adresen et al.[17] found that people who suffered pain as part of their disorder rated their pain as worse than did people who did not
have the disorder, and, in a different sort of study, Todorov and Kirchner[18] found that proxies under-reported symptoms of people with disabilities.

A second question was whether a discrepancy could result from vagueness in the judgment scale. In particular, if the scale has unclear anchors — that is, unclear standards of comparison — people with a health disorder might evaluate having the disorder by comparing themselves to others with the disorder, rather than to those without it. This possibility predicts that the discrepancy would be larger when the judgment scale is less clear.

We used three methods of eliciting judgments of undesirability, an anchored scale, a vague scale, and a happiness scale. The anchored scale is anchored at the bottom by death and at the top by the absence of the disorder being rated. The vague scale was anchored at the bottom by “extremely undesirable” and at the top by “not undesirable at all.” The happiness scale asked about the overall effect of the disorder on happiness. Both ends were vaguely described: “greatly increased happiness” and “greatly decreased happiness.”

The third main question concerns judgments of self vs. others. For each scale, subjects judged for themselves and for someone else. A Self-Other difference might help to explain past findings of a Have–Not-have discrepancy.

2.1 Method

Eighty subjects completed a questionnaire on the World Wide Web, at http://www.psych.upenn.edu/~baron/qs.html. Their ages ranged from 16 to 60 (median 34); 27.6% were male; 15% were students. Most subjects had completed other studies on other topics at the same site. They had originally discovered it either through search engines, links from other sites (such as those listing ways to earn money on the web), or “word of mouth.” In general, the population of visitors to this site, while not all U.S. residents, has about the same median education and income as the U.S. adult population. Other than the subjects being web users, their most salient characteristic is that they are mostly women. The questionnaire began:
Preference for health conditions

The following study concerns judgments of chronic health conditions. . . .

One [type of] question concerns undesirability, which means the strength of preference for not having the condition.

[Another type of] question concerns the effect of the condition on overall happiness or unhappiness.

You will make the undesirability ratings from two points of view. One is your own point of view, if you had the condition. If you have or have had it, please pay attention to the description, and think about that rather than your own case.

The other is the point of view of the average person.

In each case, imagine that either you or the average person has had the condition for 6 months. Also, the condition will not change in the foreseeable future. It will not get better, and it will not get worse. This is important. Do not suppose that it will improve.

A typical item (one screen) using the anchored scale for both Self and Other ratings is:

Item 1 out of 45:

Suppose that you had the condition: bad knee — running is painful.

For yourself, on a scale of undesirability in which 0 is 'not having' bad knee (with everything else the same) and

100 is 'imminent death', where would you put bad knee? [ ]

Now suppose that the average person had bad knee (running is painful).

Where do you think that the average person would put bad knee on the same scale?
The description for the vague scale was, “on a scale of undesirability in which 0 is 'not undesirable at all' and 100 is 'extremely undesirable'.” The description for the happiness scale was “on a scale of happiness/unhappiness in which 5 is 'greatly increased happiness from the condition', 0 is 'no change in happiness from the condition', and -5 is 'greatly decreased happiness from the condition.” (Formatting is omitted here.) Each subject used one scale type at a time, for all the disorders. The order of the three types was randomized across subjects.

With each scale type, the subject rated the following 15 health conditions, shown here with their brief descriptions:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>asthma</td>
<td>attacks of breathing difficulty — 1 per week</td>
</tr>
<tr>
<td>chronic back pain</td>
<td>running is impossible and walking difficult</td>
</tr>
<tr>
<td>bad knee</td>
<td>running is painful</td>
</tr>
<tr>
<td>inability to walk</td>
<td>requires a wheelchair</td>
</tr>
<tr>
<td>insomnia</td>
<td>2 hours less sleep than desired on most nights</td>
</tr>
<tr>
<td>migraines</td>
<td>debilitating 2 hour headaches — 2 per week</td>
</tr>
<tr>
<td>short stature</td>
<td>6 inches shorter than average for sex</td>
</tr>
<tr>
<td>excessive weight</td>
<td>50% more than normal weight</td>
</tr>
<tr>
<td>nearsightedness</td>
<td>glasses required</td>
</tr>
<tr>
<td>partial deafness</td>
<td>hearing aid required</td>
</tr>
<tr>
<td>nightmares</td>
<td>frightening dreams most nights</td>
</tr>
<tr>
<td>acne</td>
<td>pimples all over face</td>
</tr>
<tr>
<td>smoking habit</td>
<td>pack a day of cigarettes</td>
</tr>
<tr>
<td>arthritis</td>
<td>pain in hips or shoulders with any movement</td>
</tr>
<tr>
<td>heart disease</td>
<td>chest pain from walking or other activity</td>
</tr>
</tbody>
</table>

After completing all the ratings, the subjects were asked, “Which of the following conditions have you had yourself for at least a year? (You may include conditions that were more or less severe than the descriptions used here.)” They saw a list of the conditions (disorders), with a letter
before each, and they were asked to type the letters.

2.2 Results

2.2.1 Not-have–Have discrepancy

To look for a Not-have–Have discrepancy effect, we first standardized the ratings for each disorder across subjects (separately for Self and Other ratings with each of the three scales). This removed the effect of differences in the seriousness of disorders. (Otherwise, we would expect that Have would seem less serious than Not-have simply because the less serious disorders are more frequent.)

Then we eliminated disorders that were rare in our sample. Because of the standardization, rare disorders could end up with extreme $z$ scores. As it happened, four of the disorders (inability to walk, partial deafness, nightmares, and heart disease) occurred in 4 subjects or fewer, and all others occurred in 11 subjects or more, so we dropped these four health disorders from further analysis of the Not-have–Have discrepancy.

We then computed the mean of the standardized Have disorders for each subject, separately for Self and Other ratings and for the different scales, and the mean of the Not-have disorders. Table 1 summarizes the main results by scale type and type of difference (with $p$ levels based on two-tailed $t$ tests). Note that numbers refer to seriousness so that larger numbers are worse.

| Insert Table 1 |

The results showed a clear Not-have–Have discrepancy, when we combined the results from all three scales (which was possible, since all had been standardized). Have was considered less serious than Not-have. Across subjects, combining Self and Other, the mean $z$-score difference between Not-have and Have was .13 ($t_{74} = 2.23, p = .0291$, across subjects — note that somewhat different subjects are involved in different analyses because of missing data). The effect was present for both Self (difference .11, $t = 1.83, p = .0728$) and Other (difference .14, $t = 2.27, p = .0260$). It is apparent that the Self-other by Not-have–Have interaction did not appear. The effects was at least as large for Other as for Self.
Our main hypothesis was that the Not-have–Have discrepancy would be larger for the vaguer scales. This was not supported. Table 1 shows the relevant results. Although the discrepancy was significant for the anchored and happiness scale and not for the vague scale, the three scales did not differ significantly in the size of the discrepancy. Importantly, the discrepancy was found for the anchored scale.

2.2.2 Self-Other difference

To examine the Self-Other difference, we used all health disorders (since this difference was between questions presented on the same screen and thus had less error). First, we looked for an overall Self-Other difference across all three measures. To do this, we used the actual ratings (not standardized), but we multiplied the happiness ratings by \(-20\) to roughly equate the scales (because most ratings were between 0 and \(-5\)). Table 2 shows the results for each scale (again, with high numbers representing worse health).

— Insert Table 2 —

Combining the three measures, the overall difference amounted to a mean of 2.0 points on the 100 point scale, with ratings more severe for other than self. As shown in Table 2, the difference was significant overall and for the anchored scale. Again, the differences among scales in the size of the effect were not significant.

It thus appears that the Self-Other discrepancy exists as hypothesized, and it therefore may account somewhat for the Have–Not-have discrepancy when Have’s are asked about themselves and Not-have’s are asked about others.

Although we had no particular hypothesis about the interaction between scale type and Other-Self, it appears, again, that the anchored scale is most sensitive to the difference.

2.2.3 Differences among disorders

The Not-have–Have discrepancy depended on the disorder. To show this, we asked whether disorders with a higher discrepancy in one half of the subjects were also higher in the other half.
We measured the mean discrepancy for the odd-numbered subjects for each disorder (combining all three measures) and the mean for the even-numbered subjects. The two sets of means were correlated across the 15 disorders ($r = .74, p = .0013$ one tailed). Table 3 shows, in the rightmost column, the Not-have–Have discrepancy for the different disorders. Although we made an effort to find all the usable common disorders, it seems that the overall result of a positive discrepancy was an artifact of our sample, and a different sample might have yielded even a reversed discrepancy. For example, migraines seem to be worse to those who have them than to those who do not. (The rare disorders are in parentheses. While these numbers are suggestive of similar results, recall that only four subjects or fewer had each of these disorders.)

We found the same kind of consistency across scale measures for the Self-Other difference ($r = .89$). The means for the disorders are shown in Table 3. Table 3 also shows the mean severity ratings (averaging all three measures, with happiness multiplied by $-20$ before averaging).

— Insert Table 3 —

3 Experiment 2

The main result of Experiment 1 was a failure to find a larger discrepancy between Haves and Not-haves in the vague scale or in the happiness scale than in the anchored scale. The happiness scale did, however, show a slightly larger discrepancy than the other two scales. Moreover, the use of “increased happiness” and “decreased happiness” in the description of the happiness scale might have tended to reduce the tendency to think of one’s own disorder as a reference point. Such a tendency is one of the mechanisms that could lead to a smaller discrepancy for the happiness scale.

It is possible that a scale focusing more explicitly on quality of life, rather than seriousness of a health state, might show a larger discrepancy. Especially when the reference points are vague, people with a disorder may tend to think about others with the same disorder when they evaluate their quality of life.

In Experiment 2 (actually done after Experiment 3), we asked four questions, with vague vs.
speciﬁc crossed with severity (of an disorder) vs. quality of life (QOL).

3.1 Method

Ninety-nine subjects completed a questionnaire on the World Wide Web. Their ages ranged from 19 to 68 (median 36); 22.2% were male; 11% were students. The questionnaire began:

Health conditions

The following study concerns judgments of chronic health conditions. There are 40 questions (screens). Each question presents a short description of a health condition and asks one of two types of rating question.

One question concerns health. We ask you to rate the health condition on a 50-100 scale. Questions differ in how 50 and 100 are defined. Pay attention to these definitions.

The other question concerns the overall quality of life of those who have the condition. Note: You can go below 50 if you feel that a condition is worse than the definition of "50".

You will make the ratings from two points of view. One is your own point of view, supposing that you have the condition. If you have really had the condition, please pay attention to the description of the condition, which may be more or less serious than your own case.

The other point of view is that of the average person who gets the condition described. In each case, imagine that either you or the average person has had the condition for at least 6 months. Also, the condition will not change in the foreseeable future. It will not get better, and it will not get worse. Do not suppose that it will improve.

A typical question, in the QOL-speciﬁc condition, appeared as follows (with the definitions of 100 and 50 in green and red, respectively):
Suppose that you had the condition: acne — pimples all over face.

For yourself, on a scale of overall quality of life in which
100 is as good as that of someone with a meaningful job, friends, family, and good health, and
50 is as bad as that of someone who cannot walk more than 10 feet because of partial paralysis, has a dull job, and no close family or friends,
where would you put acne?

Now suppose that the average person had acne (pimples all over face).
Where do you think that the average person would put acne on the same scale?

The scale definitions for the QOL-vague condition were: “100 is a very good quality of life and 50 is a very poor quality of life.” (We used 50 to make it easier for subjects to assign number below the bottom anchor.) For the health-specific condition they were: “100 is as healthy as a 20-year-old with no health conditions and 50 is as serious as being unable to walk more than 10 feet because of partial paralysis.” and for the health-vague condition they were, “100 is very good health and 50 is very poor health.” The four conditions were blocked, and in a different random order for each subjects.

The health states were chosen on the basis of previous studies to be common. Their order randomized separately for each subject. Their definitions were:

Asthma: attacks of breathing difficulty — 1 per week
Back pain: pain from lifting objects
Insomnia: 2 hours less sleep than desired on most nights
Shortness: 6 inches shorter than average height for sex
Overweight: inability to lose excess weight
Nearsightedness: glasses required
Acne: pimples all over face
Smoking habit: uncontrollable addiction to cigarettes
Arthritis: pain in hips or shoulders with some movements
Heart disease: occasional chest pain from climbing stairs
At the end of the study, subjects again indicated which disorders they had for at least a year.

### 3.2 Results

Table 4 shows the mean ratings for the four conditions (after elimination of bad data, as we shall describe). Here, high numbers represent good health or good QOL. It is apparent that the four measures agreed closely on the relative seriousness of the health states, and the health states vary considerably in seriousness. Some subjects seemed to misunderstand the scale for some condition blocks. To assess misunderstanding, we correlated each subject’s scores in each block with the ten means based on all four conditions. We eliminated blocks when the correlation was less than .25. This resulted in deletion of 16.3% of the data. Two subjects had all their data deleted, and 58 had no data deleted. All statistical analysis is based on whatever data were available.

As is apparent in Table 4, Self ratings are consistently higher (less severe) than Other ratings, as found in Experiment 1 (mean difference of 2.01, $t_{57} = 6.77, p = 0.0000$). Specific ratings are also higher than vague ratings (mean difference 3.02, $t_{57} = 5.68, p = 0.0000$), but this is surely the result of the specific scale definitions, not a general result. Health and QOL ratings did not differ significantly. Self-Other difference was slightly greater for health than for QOL ($t_{57} = 2.55, p = 0.0135$); this too does not seem to imply any general conclusion.

Figure 1 shows the results for the Not-have vs. Have discrepancy for the ten health states and the four conditions (for Self and Other combined); positive numbers indicate that Not-have’s rate the disorder as worse than Have’s. We did not find an overall Not-have–Have discrepancy, even for Self. This was because the health states differed in the direction of this effect, as is apparent from Figure 1. The correlation across disorders between the discrepancy measures for two halves of the sample (computed as in Experiment 1) was .49 ($p = .0771$, one tailed).

Although the overall discrepancy (Have less severe than Not-have) was not significant, we note, first, that it was greater for Self (0.063) than for Other (0.030, in contrast to Experiment 1,
It is also apparent (from Figure 1) that the discrepancy, to the extent to which we find it, was no larger for QOL than for health. In fact, it was nearly significant for health alone in the Self condition (mean 0.25, $t_{69} = 1.92, p = 0.0596$) and in the opposite direction for QOL (-0.17), resulting in a significantly greater discrepancy (Have higher than Not-have) for health than for QOL ($t_{63} = 3.03, p = 0.0036$). This result starkly contradicts the hypothesis that the discrepancy would be greater for QOL.

Similarly, the discrepancy was no larger for vague than specific. Again, we found the opposite ($t_{63} = -2.23, p = 0.0291$): the Not-have–Have discrepancy was greater (higher ratings for Have) in the specific than in the vague condition.

### 4 Experiment 3

The purpose of Experiment 3 was to ask whether the discrepancy effect was the result of a focusing illusion. The idea of a focusing illusion is that, when people are asked about a difference between two disorders, they focus on the attributes of those disorders that are different, because of the way the question is asked. They thus exaggerate the magnitude of the difference. For example, people asked to compare life in California with life in the midwestern states of the U.S. focused on the weather, thus overestimating the benefits of living in California [14]. Likewise, people who do not have a disorder could focus on the attributes of life that are more affected by that disorder.

As we explained in the Introduction, a previous study [15] found no evidence for a focusing illusion as an explanation of the Have–Not-have discrepancy. The study tried to reduce such an illusion, if it existed, by calling subjects’ attention to a variety of attributes that characterize the goodness of life. Asking subjects how a disorder affected each of these dimensions did not affect their subsequent rating of the same disorder.

Ubel et al. [15] used rare conditions, such as below-the-knee amputation and paraplegia, so they were unable to make the Not-have–Have comparison in their sample (jurors). We do not know
whether they would have found a discrepancy effect, if they had done so. Also, it is possible that
subjects may have understood that the disorders did not affect all life attributes equally, yet, still,
reverted to the focusing illusion even after they were forced to consider all the attributes.

In the present experiment, we followed the basic design of Ubel et al. In particular, we first
asked for holistic ratings, then we asked for attribute-by-attribute ratings, and finally we asked for
holistic ratings again. However, we asked about the attribute-by-attribute ratings in a way that
allows us to do a rough calculation of utility based on multi-attribute utility theory [20]. We thus
call this a MAU elicitation. In particular, we asked for numerical ratings on each of seven attributes,
each with a clearly anchored endpoint. At the end of the experiment we asked for weights of these
attributes. To calculate the MAU (multi-attribute utility) of each disorder for each subject, we
multiplied each attribute rating by the weight of that attribute and then added up these products
across the seven attributes.

If the Not-have–Have discrepancy results from a focusing illusion, it should largely disappear in
the MAU ratings, because subjects were forced to rate attributes one at a time. Thus, even if the
effect of this forced variety is limited to the MAU task itself, we should be able to detect it. We do
not need to rely on transfer to the subsequent holistic rating task.

For the holistic question, we used paralysis from the waist down as the standard, rather than
death, because we thought that people might be reluctant to say that a disorder they had was
anywhere near as bad as death.

4.1 Method

Seventy-nine subjects completed a questionnaire on the World Wide Web. Their ages ranged from
18 to 74 (median 36); 33% were male; 14% were students.

The questionnaire began:

Preference for health conditions

This study concerns judgments of chronic health conditions. It has 24 screens. Please
read all of these instructions carefully.

In each question, you will see a short description of a health condition, and you will answer one of two types of questions.

One question concerns undesirability, which means the strength of preference for not having the condition. This type of question will come first, and then it will be repeated at the end. Do not worry about whether your answer is the same or not. Just try to answer accurately both times.

You answer this question on a scale where 0 means “not having the condition” and 100 means “as bad as being paralyzed from the waist down.” You can use numbers greater than 100 if necessary (but no greater than 200).

The other type of question concerns the negative effect of the condition on several domains of your life:

1. Pain and discomfort
2. Economic standard of living
3. Work
4. Love life
5. Family life other than love life
6. Spiritual life broadly defined
7. Leisure activities other than family life

Please try to interpret these descriptions so that they do not count the same effects twice. For example, if “spiritual life” includes communing with nature, do not also count this as part of “leisure”.

You answer these questions on a scale where 0 means “no negative effect” and the meaning of 100 is specified in the question. If you think that some health condition has
a positive effect, then use a negative number for your response. Remember, this question is about negative effects.

At the end, you will be asked a few additional questions.

Each question asks you to “suppose that you had” the condition. Of course, you may actually have it now. If so, please pay attention to the description, which may be more or less severe than your own case. Rate that description, not your own case.

In each case, imagine that you have had the condition for 6 months. Also, the condition will not change in the foreseeable future. It will not get better, and it will not get worse.

This is important. Do not suppose that it will improve.

We used the following disorders, which had showed a discrepancy in the hypothesized direction in Experiment 1 (except for heart disease, which was modified to be less severe).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>asthma</td>
<td>attacks of breathing difficulty — 1 per week</td>
</tr>
<tr>
<td>insomnia</td>
<td>2 hours less sleep than desired on most nights</td>
</tr>
<tr>
<td>short stature</td>
<td>6 inches shorter than average for sex</td>
</tr>
<tr>
<td>nearsightedness</td>
<td>glasses required</td>
</tr>
<tr>
<td>acne</td>
<td>pimples all over face</td>
</tr>
<tr>
<td>smoking habit</td>
<td>pack a day of cigarettes</td>
</tr>
<tr>
<td>arthritis</td>
<td>pain in hips or shoulders with any movement</td>
</tr>
<tr>
<td>heart disease</td>
<td>occasional chest pain from climbing stairs</td>
</tr>
</tbody>
</table>

For the holistic task, a typical question read.

Suppose that you had the condition: arthritis – pain in hips or shoulders with any movement. On a scale of overall undesirability in which

0 is not having arthritis (with everything else the same) and

100 is waist-down paralysis,
where would you put arthritis?

For the MAU task, a typical question read:

Suppose that you had the condition: insomnia – 2 hours less sleep than desired on most nights. Rate this condition for its negative effect on each of the following domains of your life. 0 means ‘no negative effect at all.’ 100 is defined for each question. (Use a negative number for a positive effect.)

1. Pain, fatigue, and discomfort: 0 = no effect; 100 = as bad as death.
2. Economic standard of living: 0 = no effect; 100 = dire poverty.
3. Work: 0 = no effect; 100 = unable to do any work.
4. Love life: 0 = no effect; 100 = love life nonexistent.
5. Family life: 0 = no effect; 100 = family life nonexistent.
6. Spiritual life: 0 = no effect; 100 = spiritual life nonexistent.
7. Leisure activities: 0 = no effect; 100 = activities nonexistent.

The order of health conditions was randomized separately for each subject and then fixed for the three parts of the questionnaire.

The weight elicitation, at the end of the questionnaire, read.

Now please rate each of the following on a scale where 0 represents ‘not bad at all’ and 100 represents ‘as bad as paralysis from the waist down’. When you make these ratings, think only about the thing you are rating. Try to imagine that everything else is the same. (This is hard, but do your best.)

Then subjects saw a list of all the lower ends of each scale in the MAU task, e.g, “Pain, fatigue, and discomfort as bad as death.”
4.2 Results

Seven subjects were dropped because they gave 0 responses to all or most of the holistic questions in one section (including the more serious disorders) or because they gave 0 to all the questions about weights. Also, some subjects were dropped from some analyses. In particular, subjects whose holistic ratings before and after the MAU ratings did not correlate positively with each other across disorders were dropped for analyses of the holistic ratings. When holistic ratings and MAU ratings correlated negatively, we dropped the one with the lowest correlation with the mean of all severity measures across disorders. By these criteria, we dropped one subject from analysis of the MAU data and 13 subjects from analysis of the holistic data. The first holistic task seemed particular prone to induce reversed judgments (high numbers for less serious disorders).

To compute MAU utilities, we first re-scaled the weights for each subject so that the maximum weight was 1. Likewise, we rescaled the ratings so that the highest rating given by each subject was 1. Then we multiplied the weights by the ratings for each attribute for each disorder for each subject. Notice that the weights were elicited using the same end points as those used in the rating task.

To assess the Not-have–Have discrepancy, we used the standardized responses as in Experiment 1. We also standardized the MAU utilities by disease.

The Not-have–Have discrepancy was significant overall, averaging the holistic (averaged over the two parts) and MAU ratings (mean difference .16 in terms of standardized scores, \( t_{61} = 2.19, p = .0320 \), two tailed; positive numbers mean that Have is less serious than Not-have). It was also significant for the MAU ratings (mean 0.15, \( t = 2.51, p = .0144 \)) but not for the holistic ratings (mean 0.14, \( t_{64} = 1.57, p = .1213 \)). These results are a clear rejection of the focusing hypothesis, which predicted an effect for the holistic ratings but not for the MAU ratings.

Again, the discrepancy varied across disorders, as shown in Figure 2. The split half correlation across the eight disorders in the discrepancy (computed as in Experiments 1 and 2) was 0.70 (\( p = .0257 \) one tailed, with 6 df).
The pattern of attribute ratings should vary by disorder. To test whether this was true, we performed an analysis of variance of the attribute ratings, using subject, attribute and disorder as factors. Importantly, the interaction between attribute and disorder was significant ($F_{42,3276} = 28.05, p = .0000$). Main effects of attribute and disorder were also large and significant. Table 5 shows the mean ratings, on a scale where the worse end of each attribute is 1 and “no effect” is 0, and the Not-have–Have discrepancy on the same scale. The interaction between attribute and disorder is illustrated by the high numbers for the effect of arthritis on pain and the effect of acne (and shortness) on love life.

We examined the Not-have–Have discrepancy for each attribute. The largest discrepancies were those for family life ($t_{72} = 3.18, p = .0022$, two tailed), love life ($t = 2.10, p = .0396$), and spiritual life ($t = 3.53, p = .0007$). The effect for work was nearly significant ($t = 1.75$). No other effect was significant by the usual criterion, although all were in the hypothesized direction, including pain and discomfort. These results suggest that those who do not have the disorders may fail to recognize how little effect they have on some domains.

Discussion

Our results indicate that the Have–Not-have discrepancy can be studied with in web respondents and common disorders. Although a minority of subjects had each disorder, most of them had at least one. For more general purposes the list could be expanded to include other aspects of life aside from health, such as being single, poor, unemployed — or good things such as being wealthy. Indeed, we may have done this by include shortness as a disorder. We have no reason to think that the general principles underlying the discrepancy would differ for non-health states.

We found consistent differences among disorders in the magnitude and direction of the discrepancy. One possible explanation of these differences — and there may well be more than one
is that the usual discrepancy (Not-have worse than Have) is found for disorders that have an
even. A limitation of this conclusion follows from our use of within-subject designs. It is possible that
those who have a disorder recalibrate their entire scale for judgments of all disorders (or of their
absence). If, for example, a paraplegic rated himself as “happy,” he might still believe that others
are happier and rate a typical person without paraplegia as “ecstatic.” We do not think that such a
general re-calibration is likely for the kinds of scales we used, for the kinds of minor disorders that
we studied, but such recalibration may exist elsewhere. If it exists, a different design is required to
detect it.

We also found further evidence that de-focusing manipulations do not reduce the discrepancy.
Unlike earlier studies, these results did not rely on the carry-over from a de-focusing manipulation
to another task. We found the discrepancy in the MAU task itself.

Also, if the focusing hypothesis were true, we might have found a larger discrepancy for health
than for QOL in Experiments 1 and 2 (especially Experiment 2, which was better controlled), and
we did not find this. Instead, we found a larger discrepancy for health than for quality of life in
Experiment 2.

The focusing hypothesis is not quite dead, however. It is possible that, even within a life domain,
people with a disorder focus on sub-attributes of that domain that are less affected by the disorder.
It may be difficult to draw a line between “attributes” and “activities,” however. It may therefore
be difficult to distinguish this form of the focusing hypothesis from other hypotheses that involve
knowledge of specific adaptations (such as using a computer for reading when one is blind).
Finally, people think they will adapt better than others. In Experiment 2, this is more true when they have in fact adapted (Haves). This result could explain some of the discrepancy found in previous studies, if people who do not have the disorder think about others rather than themselves.

The main remaining explanations of the discrepancy are failure of Not-haves to predict adaptation and self-deception. This is an important distinction. Adaptation is real, and self-deception is, in an important sense, false. We may think of people’s values in terms of what Keeney calls fundamental values and means values [21]. People’s good is in their fundamental values. (These may include wanting their means values honored.) Means values are connected to fundamental values through beliefs. If beliefs are false, then the means values lose their claim to represent a person’s good [22]. Thus, when we help people make decisions, we do not do them any good when we honor their values based on self-deception and when these conflict with their fundamental values. If people with disabilities are deceiving themselves about how close to normal they are, then, to this extent, we should discount their judgments and work harder to cure and prevent their disorders than their own judgments would imply.

On the other side, mis-prediction of adaptation is a false belief held by Not-haves. To the extent to which adaptation is real and not predicted, then we should discount these judgments in evaluating the seriousness of disorders. Of course, both kinds of error could be true. And our results suggest a different kind of Not-have error as well, a failure to appreciate the seriousness of disorders that have no external manifestation.
References


[5] Llewellyn-Thomas HA, Sutherland HJ, Theil EC. Do patients evaluations of a future health state change when they actually enter that state? Medical Care, 1993; 31: 1002–1012.


**Table 1:** Relationship between scale type and discrepancy, in z scores of seriousness (mean of Self and Other, common disorders only), Experiment 1.

<table>
<thead>
<tr>
<th>Scale type</th>
<th>Haves</th>
<th>Not-haves</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchored</td>
<td>−0.08</td>
<td>0.04</td>
<td>.0430</td>
</tr>
<tr>
<td>Vague</td>
<td>−0.03</td>
<td>0.02</td>
<td>n.s.</td>
</tr>
<tr>
<td>Happiness</td>
<td>−0.13</td>
<td>0.04</td>
<td>.0479</td>
</tr>
<tr>
<td>Overall</td>
<td>−0.08</td>
<td>0.03</td>
<td>.0291</td>
</tr>
</tbody>
</table>

**Table 2:** Relationship between scale type and Self-Other difference, in seriousness scores on a rough 100 point scale (higher numbers representing worse health), all disorders, Experiment 1.

<table>
<thead>
<tr>
<th>Scale type</th>
<th>Self</th>
<th>Other</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchored</td>
<td>38.7</td>
<td>41.2</td>
<td>.0028</td>
</tr>
<tr>
<td>Vague</td>
<td>55.9</td>
<td>56.8</td>
<td>n.s.</td>
</tr>
<tr>
<td>Happiness</td>
<td>44.9</td>
<td>47.6</td>
<td>.0609</td>
</tr>
<tr>
<td>Overall</td>
<td>46.5</td>
<td>48.5</td>
<td>.0105</td>
</tr>
</tbody>
</table>
Table 3: Mean severity ratings for the 11 common disorders, and Self-Other (positive for Other worse than Self) and Not-have–Have differences (positive for Not-have worse than Have), Experiment 1 (ordered by discrepancy).

<table>
<thead>
<tr>
<th>Health disorder</th>
<th>Number having</th>
<th>Mean severity (0–100 scale)</th>
<th>Self-Other difference (0–100 scale)</th>
<th>Not-have–Have difference (z scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>23</td>
<td>41.9</td>
<td>−14.0</td>
<td>0.41</td>
</tr>
<tr>
<td>Acne</td>
<td>13</td>
<td>43.6</td>
<td>4.2</td>
<td>0.39</td>
</tr>
<tr>
<td>Short</td>
<td>29</td>
<td>25.5</td>
<td>12.1</td>
<td>0.32</td>
</tr>
<tr>
<td>Insom</td>
<td>22</td>
<td>37.3</td>
<td>3.2</td>
<td>0.09</td>
</tr>
<tr>
<td>Asthma</td>
<td>11</td>
<td>49.2</td>
<td>3.2</td>
<td>0.09</td>
</tr>
<tr>
<td>Arthritis</td>
<td>18</td>
<td>56.5</td>
<td>0.8</td>
<td>0.05</td>
</tr>
<tr>
<td>Nearsgt</td>
<td>43</td>
<td>21.7</td>
<td>6.5</td>
<td>0.04</td>
</tr>
<tr>
<td>Back</td>
<td>21</td>
<td>65.1</td>
<td>−1.6</td>
<td>−0.01</td>
</tr>
<tr>
<td>Weight</td>
<td>10</td>
<td>56.0</td>
<td>0.5</td>
<td>−0.02</td>
</tr>
<tr>
<td>Knee</td>
<td>12</td>
<td>38.1</td>
<td>4.0</td>
<td>−0.11</td>
</tr>
<tr>
<td>Migr</td>
<td>22</td>
<td>56.4</td>
<td>−0.8</td>
<td>−0.15</td>
</tr>
</tbody>
</table>
Table 4: Mean ratings in Experiment 2, for ten health states in the four conditions, for self and other (based on cleaned-up data); higher numbers represent better health.

<table>
<thead>
<tr>
<th></th>
<th>Health specific</th>
<th>Health vague</th>
<th>QOL specific</th>
<th>QOL vague</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearsighted</td>
<td>92.1</td>
<td>92.3</td>
<td>91.0</td>
<td>90.1</td>
</tr>
<tr>
<td></td>
<td>89.5</td>
<td>87.9</td>
<td>88.8</td>
<td>88.0</td>
</tr>
<tr>
<td>Short</td>
<td>90.7</td>
<td>90.0</td>
<td>90.9</td>
<td>90.1</td>
</tr>
<tr>
<td></td>
<td>86.1</td>
<td>85.3</td>
<td>86.6</td>
<td>85.5</td>
</tr>
<tr>
<td>Acne</td>
<td>87.1</td>
<td>83.4</td>
<td>88.3</td>
<td>84.2</td>
</tr>
<tr>
<td></td>
<td>82.3</td>
<td>78.4</td>
<td>82.9</td>
<td>80.8</td>
</tr>
<tr>
<td>Insomnia</td>
<td>83.5</td>
<td>79.4</td>
<td>82.3</td>
<td>79.8</td>
</tr>
<tr>
<td></td>
<td>80.7</td>
<td>77.1</td>
<td>81.1</td>
<td>78.9</td>
</tr>
<tr>
<td>Back</td>
<td>81.8</td>
<td>75.6</td>
<td>77.9</td>
<td>76.4</td>
</tr>
<tr>
<td></td>
<td>78.8</td>
<td>74.6</td>
<td>75.9</td>
<td>75.5</td>
</tr>
<tr>
<td>Asthma</td>
<td>80.2</td>
<td>76.5</td>
<td>76.9</td>
<td>73.8</td>
</tr>
<tr>
<td></td>
<td>77.2</td>
<td>73.2</td>
<td>74.8</td>
<td>72.5</td>
</tr>
<tr>
<td>Overweight</td>
<td>78.3</td>
<td>77.3</td>
<td>76.8</td>
<td>72.3</td>
</tr>
<tr>
<td></td>
<td>75.2</td>
<td>73.5</td>
<td>74.9</td>
<td>70.0</td>
</tr>
<tr>
<td>Smoking</td>
<td>80.4</td>
<td>72.4</td>
<td>74.3</td>
<td>70.4</td>
</tr>
<tr>
<td></td>
<td>78.3</td>
<td>73.6</td>
<td>75.9</td>
<td>71.6</td>
</tr>
<tr>
<td>Arthritis</td>
<td>77.8</td>
<td>74.4</td>
<td>73.8</td>
<td>72.5</td>
</tr>
<tr>
<td></td>
<td>74.8</td>
<td>71.4</td>
<td>72.2</td>
<td>71.6</td>
</tr>
<tr>
<td>Heart</td>
<td>74.7</td>
<td>71.3</td>
<td>70.6</td>
<td>67.8</td>
</tr>
<tr>
<td></td>
<td>71.4</td>
<td>68.6</td>
<td>68.6</td>
<td>66.2</td>
</tr>
<tr>
<td>MEAN</td>
<td>82.7</td>
<td>79.3</td>
<td>80.3</td>
<td>77.8</td>
</tr>
<tr>
<td></td>
<td>79.4</td>
<td>76.4</td>
<td>78.2</td>
<td>76.1</td>
</tr>
</tbody>
</table>
Table 5: Mean ratings and discrepancies in these ratings as a function of attribute and disorder, Experiment 3; higher numbers represent worse health.

<table>
<thead>
<tr>
<th></th>
<th>Pain</th>
<th>Econ</th>
<th>Work</th>
<th>Love</th>
<th>Family</th>
<th>Spirit</th>
<th>Leisure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean disutility ratings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>0.49</td>
<td>0.20</td>
<td>0.31</td>
<td>0.22</td>
<td>0.18</td>
<td>0.09</td>
<td>0.42</td>
</tr>
<tr>
<td>Insomnia</td>
<td>0.39</td>
<td>0.19</td>
<td>0.41</td>
<td>0.27</td>
<td>0.30</td>
<td>0.14</td>
<td>0.37</td>
</tr>
<tr>
<td>Short</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
<td>0.32</td>
<td>0.12</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>Nearsighted</td>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
<td>0.15</td>
</tr>
<tr>
<td>Acne</td>
<td>0.22</td>
<td>0.15</td>
<td>0.21</td>
<td>0.59</td>
<td>0.19</td>
<td>0.09</td>
<td>0.24</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.33</td>
<td>0.37</td>
<td>0.21</td>
<td>0.31</td>
<td>0.28</td>
<td>0.16</td>
<td>0.29</td>
</tr>
<tr>
<td>Arthritis</td>
<td>0.74</td>
<td>0.41</td>
<td>0.60</td>
<td>0.44</td>
<td>0.31</td>
<td>0.11</td>
<td>0.62</td>
</tr>
<tr>
<td>Heart</td>
<td>0.64</td>
<td>0.29</td>
<td>0.48</td>
<td>0.38</td>
<td>0.30</td>
<td>0.12</td>
<td>0.52</td>
</tr>
<tr>
<td>Mean Not-have–Have discrepancy in disutility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>0.08</td>
<td>-0.02</td>
<td>0.06</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>Insomnia</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.02</td>
<td>0.17</td>
<td>-0.04</td>
<td>-0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>Short</td>
<td>0.03</td>
<td>0.09</td>
<td>-0.07</td>
<td>0.00</td>
<td>0.11</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>Nearsighted</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>-0.04</td>
</tr>
<tr>
<td>Acne</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.08</td>
<td>-0.01</td>
<td>0.03</td>
<td>0.10</td>
<td>0.10</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td>Arthritis</td>
<td>0.08</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.08</td>
<td>-0.02</td>
<td>0.06</td>
<td>-0.14</td>
</tr>
<tr>
<td>Heart</td>
<td>0.06</td>
<td>0.11</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.12</td>
</tr>
</tbody>
</table>
Figure 1: Have–Not-have discrepancy by disorder and method, Experiment 2. Positive numbers indicate that Not-have’s rate the disorder as worse than Have’s.
Figure 2: Have–Not-have discrepancy by disorder and method, Experiment 3. Positive numbers indicate that Not-have’s rate the disorder as worse than Have’s.