



University of Southern Denmark

The answer depends on pragmatic norms, semantic context-sensitivity, and epistemic reflection. A linguistic and epistemological analysis of the Danish Short Form 36 Health Survey (SF-36)

Petersen, Esben Nedenskov; Nørgaard, Birgitte

Published in:
Health Services and Outcomes Research Methodology

DOI:
10.1007/s10742-022-00272-0

Publication date:
2022

Document version:
Accepted manuscript

Citation for polished version (APA):
Petersen, E. N., & Nørgaard, B. (2022). The answer depends on pragmatic norms, semantic context-sensitivity, and epistemic reflection. A linguistic and epistemological analysis of the Danish Short Form 36 Health Survey (SF-36). *Health Services and Outcomes Research Methodology*, 22(4), 479-492. <https://doi.org/10.1007/s10742-022-00272-0>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Title

The answer depends on pragmatic norms, semantic context-sensitivity, and epistemic reflection. A linguistic and epistemological analysis of the Danish Short Form 36 Health Survey (SF-36)

Authors

Esben Nedenskov Petersen¹⁾, Associate Professor (ORCID: 0000-0003-2213-8812);

Birgitte Nørgaard²⁾, Associate Professor (ORCID: 0000-0002-2284-2335)

¹⁾ Department for the Study of Culture, University of Southern Denmark

²⁾ Department of Public Health, University of Southern Denmark

Corresponding author/reprint requests:

Birgitte Nørgaard, University of Southern Denmark, J.B. Winsløvs Vej 9B, 5000 Odense C,
Denmark

Cell phone: +45 24222613 / Email: binorgaard@health.sdu.dk

Funding

This study has not obtained any funding.

Conflicts of interest/Competing interests

Both authors declare that they have no conflicts of interests and no competing interests.

Availability of data and material

Not applicable

Code availability (software application or custom code)

Not applicable

Author contribution (according to the ICMJE recommendations)

Both authors have contributed substantially to this paper, including:

1. Substantial contributions to the conception of the work, analysis, and interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval

A this manuscript concerns a theoretical analysis, ethics approval was not necessary

Consent to participate

Not applicable

Consent for publication

Not applicable

Keywords

Gricean pragmatics; epistemology; context-sensitivity; Short Form 36; SF-36

Word count (excl. Title page, abstract, summary and references)

5393

Abstract

The SF-36 is a commonly used tool for measuring health status in a general population. Despite the overall moderate to high validity scores, certain communicative dynamics of the questionnaire deserve attention. Our aim was to examine how pragmatic dynamics, semantic context-sensitivity and epistemic reflection may influence answers to the SF-36.

We applied a three-step Gricean analysis, including identification of the items in which pragmatic dynamics are most likely to have a significant effect, examination of how Gricean maxims might affect the answers given to the items identified, and assessment of whether the combined influence of linguistic context-sensitivity and pragmatic norms is benign.

We found that the pragmatic dynamics of scalar implicatures are crucial to the interpretation of answer options but generally benign to its purpose. Regarding context, we raised concerns about the answer option ‘Ved ikke’ (Don’t know); rather than representing a neutral midpoint, the answer is compatible with both a positive and a negative answer option.

Whereas scalar implicatures are helpful to the purpose of SF-36, other contextual effects appear to be more worrisome. However, since pragmatic norms of communication, semantic context-sensitivity, and attention to epistemic error possibilities can all be expected to shape answers to the SF-36, we think that all three factors belong in a description of how the questionnaire works.

1. Introduction

Patient-reported outcomes (PRO data) are increasingly acknowledged in health care as valid data and used for both individualized and tailored health care delivery and for developing health care delivery on an organizational level. In short, PRO data are collected using either disease-specific or generic instruments (Weldring and Smith, 2013). The SF-36 (Short Form) is an example of a widely and commonly used generic tool for measuring health status (functional health and well-being) in a general population. The original SF-36 was derived from a longer instrument developed in the Medical Outcome Study (MOS) conducted by the RAND Corporation.¹ The questionnaire is available in a public license-free form (the 1st version), and a shorter version containing 12 items is also available (Hays and Shapiro, 1992).

The SF-36 was designed for self-reporting or for use by an interviewer conducting face-to-face or telephone interviews using a standardized script. Completing the questionnaire takes 5–10 minutes. The questionnaire is intended for adults – some versions specify anyone over the age of 14 – but it might provide a lower response rate for a population above the age of 65 years old (Ware and Sherbourne, 1992) or a population with a lower level of education (Bjorner et al., 1998a). The instrument was designed for use in clinical practice (screening individual patients), research (differentiating outcomes based on different treatments) and health policy evaluation (comparing the burden of different treatments), as well as for monitoring both specific and general populations (Aaronson et al., 1992, Ware and Sherbourne, 1992). The SF-36 has been translated and adapted in 29 countries and validated in numerous different patient groups, socioeconomic situations and diagnoses (Ware et al., 1996).

¹ https://www.rand.org/health/surveys_tools/mos/mos_core_36item.html

The instrument includes 36 items distributed across eight domains: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions. Furthermore, it includes a single item assessing perceived change in health. Responses are given on Likert scales (three, five or six points) or by yes/no options, and higher scores indicate a more favorable health state (Cech and Martin, 2012). The eight health profiles are derived from summarized scores, and all dimensions are independent of each other. A comprehensive manual and interpretation guide is available (Ware et al., 1996, Ware and Sherbourne, 1992).

In general, the SF-36 produces results of moderate to high validity regarding content, construct and criterion validity – also compared to other generic health instruments (McHorney et al., 1993, Scott et al., 1999). The SF-36 also produces results of reliability and sensitivity (Anderson et al., 1996), and finally, of responsiveness (Brown et al., 2009). For decades, the SF-36 questionnaire has been commonly used as a generic tool to provide a brief general measure of health-related quality of life internationally (Bjorner et al., 1998a, Guilfoyle et al., 2009), as well as in a Danish setting (Bjorner et al., 1998b). The Danish version was published and validated in 1998; however, the translation was not without challenges (Bjorner et al., 1998b). Overall, the agreement among the three translators regarding the items was low, with an intraclass coefficient of .29, whereas the translation of the response categories was considered easy, with a quality rating of 100 for clarity, common language use and conceptual equivalence (Bjorner et al., 1998b).

Despite the overall moderate to high validity scores of the SF-36 and a kind of unquestioned confidence in its nearly universal understandability and seamless applicability, we

believe that certain communicative dynamics of the questionnaire deserve more careful attention. These aspects of the SF-36 concern the kinds of pragmatic norms that structure ordinary linguistic communication (Grice, 1989, Horn, 1972, Horn, 2006) and semantic context-sensitivity (Recanati, 2010, Carston, 2002, Stanley, 2000). We also believe that the small group of questionnaire items with a ‘Don’t know’ answer option merit attention because of how epistemic reflection may give rise to context effects that might influence how participants answer these items (Ichikawa and Steup, 2018, Lewis, 1996).

Our aim in this study was to examine how pragmatic dynamics, semantic context-sensitivity and epistemic reflection may influence answers to the SF-36. Rather than focusing on a single parameter that might affect responses, we considered the possible effects of conversational norms, semantic context-sensitivity, and epistemic reflection together, in order to understand how these aspects of communication affect the surface validity of the questionnaire and answers of respondents.

2. Methods

Linguistic research building on the seminal work of Grice (Grice, 1989) models linguistic communication as a cooperative endeavour in which participants comply with an overarching Cooperative Principle (CP) and expect their interlocutors to do so as well. CP enjoins a speaker to ‘make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged’ (Grice, 1989). According to Grice, observing this principle corresponds to the observance of a number of more specific

maxims:

QUALITY: Try to make your contribution one that is true.

1. Do not say what you believe to be false.
2. Do not say that for which you lack evidence.

QUANTITY:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

RELATION: Be relevant.

MANNER: Be perspicuous.

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief. (Avoid unnecessary prolixity.)
4. Be orderly.

Following Gricean or neo-Gricean approaches (Huang, 2017), a standing expectation that speakers' respect these (or a similar set of) norms (Horn, 2004, Levinson, 2000) greatly shapes our interpretations of utterances, and obedience to these norms strongly affects what speakers deem appropriate to say. For example, if someone looking for the chocolate asks you where the

chocolate is, and you know that it was in the cupboard last night and that it is currently in the refrigerator, obedience to CP will lead you to answer only that ‘The chocolate is in the refrigerator’ because this is the most helpful answer given the purpose of the conversation. What you definitely will not say, if you are being cooperative, is ‘Last night the chocolate was in the cupboard’ because this will lead your interlocutor to think that this is the most relevant information you have to convey given what you know (Hansen et al., 2017).

Conversational norms have been shown to influence answers to questionnaires in various ways (Clark and Schober, 1992, Schwarz, 1996). Pragmatic dynamics plausibly affect how ambiguous questions are disambiguated (Schwarz, 1995) and how answer options are interpreted when ordinary language terms for ratings are combined with numerical values for ratings (Schwarz and Hippler, 1991). The connection between ambiguity and familiar concerns about language imprecision (Pepper, 1981) suggests that the effects of pragmatic dynamics on questionnaires are extensive. Together with the pervasive context-sensitivity of ordinary language meaning (Recanati, 2010, Carston, 2002, Stanley, 2000), they plausibly have a significant impact on questionnaire interpretation, as suggested by experimental data concerning vague quantifiers such as ‘rarely’, ‘often’ and ‘quite a bit’ (Wänke, 2002, Wright et al., 1994). Another potential context effect relates to contextual variance in the willingness to ascribe knowledge. Certain aspects of epistemic reflection have the potential to influence the tendency of respondents to opt for ‘Don’t know’ answers when completing the SF-36 (Lewis, 1996, Nagel et al., 2013, Alexander et al., 2014, Buckwalter, 2014).

Questions about surface validity and sense-masking may also be investigated by means of cognitive interviews (Mallinson, 2002, Collins, 2003, Conrad and Blair, 2004) aiming to uncover the reasoning processes through which respondents settle on their answers to a

survey. But while we do not dispute the value that such interviews have in survey methodology, the struggle with various limitations. Whether from probing interviews or think aloud approaches verbal reports are poorly suited to investigate those aspects of sense-making that are outside conscious access (Collins 2003, Conrad and Blair. 2004). Pragmatic and lexical analysis of response options may help to mitigate this shortcoming by identifying how pragmatic norms and semantic context-sensitivity may affect questionnaire answers, although their influence is not consciously recognized.

To examine where such communicative dynamics might influence how respondents interpret and respond to the SF-36, we carried out a three-step analysis. The first part of the analysis identified the items where pragmatic dynamics were most likely to have a significant effect. To that end, the questionnaire was examined for items where the conventional lexical meaning renders affirmation of one or more answer possibilities compatible with affirmation of one or more other answer possibilities in the same item. In such cases, the semantic content of the answer possibilities will be insufficient for some respondents to identify a unique true answer among the answer options. This, in turn, increases the probability that the task of identifying the most suitable answer, as instructed by the preamble of the SF-36, is solved by relying on pragmatic inference.

The second part of the analysis examined how Gricean maxims might affect the answers given by respondents to the items identified in the first part of the analysis. The purpose of this part was to specify the exact maxims likely to influence responses in order to help determine the overall impact of pragmatic dynamics on the validity of the questionnaire. The third part of the analysis built on the second part to assess whether the combined influence of linguistic context-sensitivity and pragmatic norms on answer choice is benign or negatively

affects the validity of the SF-36. In this part of the analysis, the basic Gricean framework was supplemented with recent work in experimental epistemology, indicating that knowledge ascriptions are affected by a subject's awareness of salient error possibilities relative to a belief. These theoretical insights were employed to assess how the answer option 'Ved ikke' (Eng. 'Don't know') may impact the answer choice depending on the circumstances of the respondent.

Because our analysis depends on the specific meanings of the Danish expressions in the Danish version of the SF-36, the English translations we give are translations of the wording in the Danish questionnaire. These translations from the Danish do not necessarily correspond to the wording of the items in the original English-language version of the SF-36. The purpose of providing the approximate translations of the items in Danish is to allow non-Danish speakers to follow the analysis where it depends on the lexical meanings of Danish words. For the English wording of the SF-36, we refer interested readers to the original questionnaire in English.

The analysis focuses on the answer options or response category labels rather than items or item stems in order to examine how the choice between different answer options might look from the point of view of a respondent aiming to communicate co-operatively. For the same reason, the analysis of answer options considers the options in relation to specific items because this is how they are encountered by a respondent. In addition, since each response label is elliptic in the sense that its communicative content depends on the item under which it occurs, an analysis focusing exclusively on response labels would be inadequate to represent the meaning communicated by picking a specific answer to an item.

To apply the Gricean framework for the pragmatic analysis the authors first analyzed the ordinary language meanings of answer options to determine the conventional

meaning of each answer option and to identify semantic context-sensitivity. Linguistic work on degree semantics for gradable predicates (Seuren, 1978, Kennedy, 1999, Kennedy, 2001, Kennedy and McNally, 2005) provided the main theoretical basis for this part of the analysis. Both this analysis of lexical meaning and the subsequent application of the Gricean framework were done collectively by the authors. Questions about the lexical meanings of terms were settled by inference from standard assumptions about similar expressions. There was no subsequent disagreement about the pragmatic analysis.

3. Results

The items identified as having answer options such that more than one answer option might be true simultaneously for a respondent are the following: 1, 2, 6, 7, 8, 9a-i, 10, 11a-d. To best capture how lexical meaning, context-sensitivity, conversational norms, and epistemic reflection are connected in our analysis, the presentation of its results focuses on a subset of these items, namely items 6, 9a-i, and 11a-d. A brief subsequent presentation details the findings regarding items 1, 2, 7, and 8.

We report our results concerning items 6, 9a-i, and 11a-d in two parts. Each part presents a detailed analysis of a specific item from the SF-36 and identifies further items to which the analysis applies because of structural similarities between the items. The analyses each follow the three steps of the method outlined in section 2.

Our first findings concern item 6 of the Danish SF-36 (corresponding to item 20 in the English version). The item reads: ‘Inden for de sidste 4 uger hvor meget har dit fysiske

helbred eller følelsesmæssige problemer vanskeliggjort din kontakt med familie, venner, naboer eller andre?’ (Eng. ‘During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours or groups?’). The answer options are ‘Slet ikke’, ‘Lidt’, ‘Noget’, ‘En hel del’, ‘Virkelig meget’ (Eng. ‘Not at all’, ‘A little’, ‘Some’, ‘Quite a bit’, ‘Very much’). These options are ordered so that the fifth option is compatible with the two preceding options, the fourth is compatible with the preceding option, the third is compatible with the second, fourth and fifth options, and the second is compatible with the third option. So, unless the first answer option, ‘Slet ikke’(Eng. ‘Not at all’), is true of the participant, more than one of the remaining answers will be true. Lexically, the pronoun ‘Noget’ (Eng. ‘Some’). denotes any value on the positive range of the degree scale and does not entail any alternative answer option. The pronoun ‘Lidt’ (Eng. ‘A little’). entails ‘Noget’ (Eng. ‘Some’). but does not entail any other answer option, while the lexicalized meaning of ‘En hel del’ (Eng. ‘Quite a bit’). semantically entails the preceding options but not ‘Virkelig meget’ (Eng. ‘Very much’). Furthermore, because of its compositional lexical meaning, ‘Virkelig meget’ (Eng. ‘Very much’). entails the two preceding options by denoting the upper end of a scale measuring amounts.

The fact that several answer options are compatible and available, however, does not entail that the choice between them is unclear. Pragmatic dynamics may supplement the participant’s interpretation of the answer options. In this case, such dynamics are likely to arise from the options available to the participant. Because of their links to scales, the answer options give rise to scalar implicatures (Horn, 1972, Horn, 2006, Israel, 2004), which aid the interpretation of the logically compatible answer options. To be as informative as the questionnaire requires, the participant has to pick the most informative option she can without

violating the Quality maxims. Thus, if she is in a position to answer ‘Virkelig meget’ (Eng. ‘Very much’), opting for another answer will violate Quantity 1 by being less informative than the questionnaire allows. If Quality permits her to answer ‘En hel del’ (Eng. ‘Quite a bit’), but not ‘Virkelig meget’ (Eng. ‘Very much’), then answering ‘Noget’ will be less informative than required by CP.

In addition, because ‘Noget (Eng. ‘Some’)’ is logically compatible with the whole range of positive degrees, ‘Noget’ will be interpreted as communicating that ‘Lidt’, ‘En hel del’ and ‘Virkelig meget’ (Eng. ‘A little’, ‘Quite a bit’, and ‘Very much’). all misrepresent the participant’s judgement. This may represent the participant’s inability to determine the answer to the item by anything more precise than an entirely unspecified non-zero value. Or it may indicate that the answer options ‘Lidt’, ‘En hel del’ and ‘Virkelig meget’ (Eng. ‘A little’, ‘Quite a bit’, and ‘Very much’). are not suitable to represent the degree to which contact has been made difficult for the participant by her health issues. Of these two possibilities, ‘Noget’ (Eng. ‘Some’). will likely be interpreted as communicating the latter, given how ‘Lidt’ and ‘En hel del’ (Eng. ‘A little’ and ‘Quite a bit’). are interpreted. ‘Lidt’ has a lexicalized link to a scale that orders a continuum of amounts with a zero amount as the upper bound and an infinitely little, non-zero amount as the maximum value. This connection to a scale of littleness gives ‘Lidt’ the interpretation ‘(at least as little as) a little’ . Correspondingly, a lexically defined relation to a reverse scale on the same continuum, going towards an infinitely great amount, gives ‘En hel del’ the interpretation ‘(as least as much as) quite a bit’ and ‘Virkelig meget’ the interpretation ‘(at least as much as) very much’ . Consequently, the answer option ‘Noget’ (Eng. ‘Some’). will tend to be interpreted as representing amounts that are not among the values ‘(at least as little as/no more than) a little’ and ‘(at least as much as/no less than) quite a bit’ .

With respect to the first question of item 6 in the Danish SF-36, the pragmatic dynamics hence seem benign with respect to the aim of interpreting answers as they are intended by participants. There is a caveat, however. Because *Lidt*, *‘En hel del’* and *‘Virkelig meget’* (Eng. *‘A little’*, *‘Quite a bit’*, and *‘Very much’*) are context-sensitive expressions, which depend on a speaker’s context-of-utterance to determine their values (Recanati, 2010, Carston, 2002, Stanley, 2000, Kennedy and McNally, 2005), the distances between them remain less than entirely clear, as does their comparative span. It is unclear, for example, how the length of a segment ranging from *‘Lidt’* (Eng. *‘A little’*) to zero (the upper bound for maximal littleness) compares to the length of the segment corresponding to more than *‘Lidt’* (Eng. *‘A little’*) but less than *‘En hel del’* (Eng. *‘Quite a bit’*). This, in turn, leaves it unclear whether the different answer options provide equally accurate means of representing judgements along the continuum of amounts. Furthermore, because their values shift between different contexts, the precise scalar degree they represent might be different for different participants on different occasions. Although their lexically encoded meanings fix the ordering of *‘Lidt’*, *‘En hel del’* and *‘Virkelig meget’*, the encoded meanings fall short of determining which specific scale segments they each denote.

Owing to structural similarities, these results carry over to the pragmatic dynamics influencing the interpretation of other answer options. The third question under item 6 (corresponding to item 22 in the English version) has exactly the same answer options as the first question. There are also sufficient structural similarities to conclude that the pragmatic dynamics work similarly for items 9a–i (items 23–31 in the English version) and 10 (item 32 in the English version), where the answer options are *‘Hele tiden’* (Eng. *‘All of the time’*), *‘Det meste af tiden’* (Eng. *‘Most of the time’*), *‘En hel del af tiden’* (Eng. *‘A good bit of the time’*), *‘Noget af tiden’*

(Eng. ‘Some of the time’), ‘Lidt af tiden’ (Eng. ‘A little of the time’) and ‘På intet tidspunkt’ (Eng. ‘At no time’). For these items, we would also expect the pragmatic dynamics to be benign with respect to ordering but problematic with respect to how the segments of the answer scale are divided because the terms of the answer options have context-sensitive semantic values.

The items 11a–d (items 33–36 in the English version) call for a different analysis. To examine their pragmatic dynamics, we consider 11b (item 34). In this item, the participant is asked to respond to the statement ‘Jeg er lige så rask som enhver anden, jeg kender’ (Eng. ‘I am as healthy as anybody I know’) with one of the following five answer options: ‘Helt rigtigt’ (Eng. ‘Completely correct’), ‘Overvejende rigtigt’ (Eng. ‘Predominantly correct’), ‘Ved ikke’ (Eng. ‘Don’t know’), ‘Overvejende forkert’ (Eng. ‘Predominantly wrong’) and ‘Helt forkert’ (Eng. ‘Completely wrong’). Here ‘Helt rigtigt’ entails ‘Overvejende rigtigt’, while ‘Overvejende rigtigt’ is compatible with ‘Helt rigtigt’ but does not entail it. Both ‘Helt rigtigt’ and ‘Overvejende rigtigt’ are incompatible with both ‘Helt forkert’ and ‘Overvejende forkert’. Similarly, ‘Helt forkert’ entails ‘Overvejende forkert’, while ‘Overvejende forkert’ is compatible with but does not entail ‘Helt forkert’. ‘Ved ikke’ (Eng. ‘Don’t know’) is compatible with all of the other four answer options, entails none and is entailed by none. That a participant does not know whether her health is excellent does not entail that it is not, nor that it is, because the factors that determine one’s health are fairly (albeit, perhaps not entirely) independent of how one makes assessments about one’s health. There is no connection between a person’s health and their knowledge of their health that ensures that a respondent knows how healthy she is.

The relations of the answer options in item 11d to scales measuring degrees of correctness ensure that their interpretation is strongly influenced by scalar implicatures. Both Quality and Quantity contribute to the pragmatic dynamics responsible for these implicatures.

Although a participant in a position to answer ‘Helt rigtigt’ (Eng. ‘Completely correct’) without violating Quality will also be in a position to answer ‘Overvejende rigtigt’ (Eng. ‘predominantly correct’) without violating Quality, Quantity would require her to answer ‘Helt rigtigt’ for her answer to be appropriately informative. If a participant opts to answer ‘Overvejende rigtigt’, she thereby communicates that she is not in a position to answer ‘Helt rigtigt’ without violating Quality. The relations between ‘Helt forkert’ (Eng. ‘Completely wrong’) and ‘Overvejende rigtigt’ (Eng. ‘Predominantly correct’) are similar. Answering ‘Overvejende rigtigt’ will implicate that the participant is not in a position to choose the answer option ‘Helt rigtigt’, whereas ‘Helt rigtigt’ will be the answer required from the participant for compliance with CP whenever a participant is in a position to answer ‘Helt rigtigt’ without violating Quality.

The interpretation of ‘Ved ikke’ (Eng. ‘Don’t know’) (present only in items 33–36) is strongly influenced by how the other answer options are interpreted. ‘Helt rigtigt’ and ‘Overvejende rigtigt’ (Eng. ‘Completely correct’ and ‘Predominantly correct’) are related semantically to an ordinal scale that orders a continuum of degrees of correctness (assumed by the Danish questionnaire) with completely correct as the maximum value. ‘Helt forkert’ and ‘Overvejende forkert’ (Eng. ‘Completely wrong’ and ‘Predominantly wrong’) have a similar link to the reverse scale (i.e., the bipolar structure of a Likert scale) with completely wrong as the maximum value. ‘Overvejende rigtigt’ hence gets the interpretation ‘(at least as correct as) Predominantly correct’, while ‘Overvejende forkert’ gets the interpretation ‘(at least as wrong as) Predominantly wrong’. So, the only unoccupied segment of the continuum of relative correctness/wrongness that ‘Ved ikke’ (Eng. ‘Don’t know’) may represent is the point with no asymmetry between relative correctness and relative wrongness.

The answer 'Ved ikke' (Eng. 'Don't know') might also relate to different ways in which the participant fails to know any of the other answer options. Assuming the standard view that knowledge is justified, true belief that is not true merely by a fortunate coincidence (Ichikawa and Steup, 2018), the most likely reasons for a respondent to judge that she does not know any answer option are that she does not believe any answer option (possibly because she is unable to adjudicate the question) or that, for each available answer option, the respondent's evidence for the answer is insufficiently justified for her to self-ascribe knowledge that the answer is right.

The latter of these options is the more troubling because there are situations in which even a subject with good evidence for a belief might be reluctant to self-ascribe knowledge. There might be situations in which a participant believes an answer option that is favored by her evidence but still thinks that she does not know the answer to be true because she considers her evidence insufficient for knowledge. In combination with the influence that awareness of salient error possibilities has on knowledge ascription, this has the potential to pressure some participants towards answering 'Ved ikke' (Eng. 'Don't know') even though their evidence favors another option. Research in experimental epistemology indicates that subjects are less inclined to self-ascribe knowledge when they become aware of ways in which their beliefs might seem true despite being false (Buckwalter, 2014, Nagel et al., 2013, Lewis, 1996). This tendency might affect the answers to 11d (item 36) from respondents who are aware of having an increased risk of undetected serious illness. A cancer survivor who fears an as yet undetected relapse, for example, would be likely to have this possibility in mind, and the same might apply to respondents who know themselves to be at an increased risk of developing diseases that are asymptomatic in their initial stages. Because the possibility of being ill without

any indications of illness will be highly salient to respondents in these situations, well-documented patterns of knowledge ascription predict that they will have a higher inclination to answer ‘Ved ikke’ to 11d (item 36) than would other respondents.

The answer options from 11d are also used in 11a–c (items 33–35), but in two of those items – 11a (item 33) and 11c (item 35) – the above concern is mitigated by the fact that the target statement assessed by a respondent explicitly focuses on the respondent’s beliefs. In 11a, the target statement is ‘Jeg bliver nok lidt lettere syg end andre’ (Eng. ‘I seem to get sick a little easier than other people’); in 11c, the statement is ‘Jeg forventer, at mit helbred bliver dårligere’ (Eng. ‘I expect my health to get worse’). As expressed by the words ‘nok’ (Eng. ‘seem’), which indicates conjecture or estimation, and ‘forventer’, which means ‘expects’, these questions are explicitly concerned with how the respondent assesses the target statement. Assuming that a respondent has relatively direct cognitive access to her assessment of such a statement, the contents of 11a (item 33) and 11c (item 35) counteract the kind of concern related to 11d (item 36).

Scalar implicatures based on Quantity should also be expected to help respondents decide their answer to 1, 2, 7, and 8, because entailments between different answer options are the source of every compatibility between answer options in the respective items. Accordingly, the effect of pragmatic norms appears benign in these cases, whereas the semantic context-sensitivity of competing answer options may render the interpretation of individual answer options sensitive to various features of the communicative context. Meanwhile, since none of these items include a reference to knowledge the peculiar contextual effects related with knowledge ascriptions should not be predicted to arise.

4. Discussion

Our Gricean analysis of the SF-36 offers important insights into the impact of pragmatic norms on a solidly validated and widely used questionnaire. The mechanisms of scalar implicature that are familiar from linguistic work on pragmatics help respondents choose between mutually compatible answer options in the SF-36 in a predictable way and hence remove some of the instrument's imprecisions. Despite the importance of scales, scalar order and terms for scale segments or points in survey research, the theoretical significance of scalar implicatures appears to have gone unnoticed in the extant literature on survey methodology.

Although previous research concerning conversational norms and survey methodology has focused on pragmatic dynamics as potential obstacles to survey research (Schwarz, 1996), the present study shows that pragmatic effects need not be harmful to the validity of reported answers. How exactly these implicatures are derived by speakers remains a hotly debated issue, but this does not jeopardize the conclusion that the pragmatic effects identified are benign. Rather than viewing pragmatic effects as inherently problematic, they should be seen as elements of a questionnaire that may be benign as well as harmful to the communicative precision of a questionnaire. Understanding how they contribute to communication is indispensable to understanding how a questionnaire such as the SF-36 works.

Such questions relating to surface validity and sense-masking may also be investigated by means of cognitive interviews (Conrad and Blair, 2004, Collins, 2003, Mallinson, 2002) aiming to uncover the reasoning processes through which respondents settle on their answers to a survey. But while we do not dispute the value that such interviews have in survey methodology, they struggle with various limitations. Whether verbal reports come from probing interviews or think aloud interviews they are poorly suited to reveal those aspects of

sense-making that are outside conscious access (Collins 2003, Conrad and Blair. 2004).

Pragmatic and lexical analysis of response options may help to mitigate this shortcoming by identifying how pragmatic norms and semantic context-sensitivity may affect questionnaire answers, although their influence is not consciously recognized.

As cognitive interviews have shown, context is crucial to how respondents answer the SF-36 (Mallinson 2002). Our study extends and refines this observation regarding the broad significance of context by contributing to a more detailed understanding of how contextual linguistic dynamics affect sense-making. Because of their context-sensitivity, terms for quantities such as 'Lidt' (Eng. 'A little') and 'En hel del' (Eng. 'Quite a bit') have their exact meaning determined relative to a context of utterance in ways that may be difficult to control. Existing research on survey methodology affirms that contextual factors may influence the understanding of survey questions by prompting respondents to invoke particular comparison classes when making judgements about typical frequency (Wänke, 2002, Wright et al., 1994). Prompting respondents to construct a comparison class based on indications about the intended target group of a questionnaire, for example, has been shown to affect responses concerning frequencies (Wänke, 2002). From a linguistic perspective, however, the main problem is that a respondent's construction of a comparison class, and the corresponding interpretations of scalar terms, are not predictable from the content of the SF-36 because there is nothing except the shifting features of different answer contexts to give them a precise meaning (Recanati, 2010, Carston, 2002, Stanley, 2000, **Kennedy and McNally, 2005**).

Our study shows that the context-sensitivity of frequency expression also affects the interpretation of answer choices with wording that is not context-sensitive. Although the meaning of 'Noget af tiden' (Eng. 'Some of the time'), for example, should not be considered

context-sensitive because it simply means a quantity of time greater than zero, the response indicated by the answer option ‘Noget af tiden’ is rendered context-sensitive by virtue of how scalar implicatures affect what it communicates. The scalar implicatures that make this answer communicate the equivalent of more than ‘Lidt af tiden’ (Eng. ‘A little of the time’ and less than ‘En hel del af tiden’ (Eng. ‘A good bit of the time’)) ensure that ‘Noget af tiden’ inherits the context-sensitivity of these adjacent answer options.

The study also raises a concern about the answer option ‘Ved ikke’ (Eng. ‘Don’t know’) in two particular questionnaire items. Rather than representing something akin to a neutral midpoint between holding it predominantly correct and holding it predominantly wrong that one’s health is excellent, the answer ‘Ved ikke’ is compatible both with a firm belief that one’s health is excellent and the belief that it is not. One problem with such compatibility between ‘Ved ikke’ and other answer options is that subjects have a higher tendency to self-ascribe ignorance when they are aware of error possibilities that describe how the evidence for a belief is compatible with the belief’s being false (Wright et al., 1994, Lewis, 1996, Nagel et al., 2013, Alexander et al., 2014). Because the salience of such error possibilities relative to the belief that one is in excellent health may be systematically higher for certain groups with specific medical histories, the effect of salient error possibilities on knowledge ascription threatens to systematically skew answers from some groups of respondents by increasing the proportion of ‘Ved ikke’ answers among them. Rather than reflecting the absence of strong belief or strong evidence, answering ‘Ved ikke’ may express epistemic caution or an increased awareness of epistemic pitfalls that other respondents might ignore. This potential effect of thinking about error possibilities raises a methodological concern not included in standard discussion of ‘Don’t know’ answers in survey methodology research (Krosnick and Presser, 2010). By drawing on

research in linguistics and epistemology, our analysis hence brings novel perspectives to bear on the issue of survey methodology.

4.1 Limitations

Although some of the dynamics that we describe may also be expected to affect versions of the SF-36 in languages other than Danish, our conclusions here are limited to the Danish questionnaire. Because the effects of pragmatic norms and epistemic reflection that we identify depend on the specific meanings of the wordings in the Danish questionnaire, separate analyses of questionnaires in other languages would be necessary to determine whether and how they are affected by such factors. Hence, it remains an open question how such factors impact the comparability of responses to the SF-36 across different languages.

We acknowledge that our study and linguistic analysis is carried out separately from practice. Meanwhile, we consider linguistic analyses important to gaining insight into unconscious reflections, and we find that our analysis offers useful insights that cannot be revealed even using cognitive interviewing.

Furthermore, although our analysis describes significant aspects of how the Danish SF-36 communicates, it does not deliver any recommendations for amendments to the questionnaire. Because there is no comparison with the validity of questionnaires with alternative wordings, the analysis cannot determine whether validity would be improved by adjusting the questionnaire. Rather than suggesting an alternative to the existing Danish questionnaire, the analysis helps to improve our understanding of the answers elicited by the questionnaire in its current form and may suggest new areas to explore by means of empirical studies including e.g., differential item functioning analyses and examination of criterion validity.

5. Conclusion

Our analysis has examined how pragmatic norms and epistemic reflection may affect the answers given by respondents to the Danish version of the SF-36. The analysis has shown that the interpretations of answer options in several questionnaire items are influenced strongly by pragmatic norms. Among the possible effects of such norms, scalar implicatures have been shown to be particularly important for the communication of the questionnaire. By giving rise to scalar implicatures, pragmatic norms help respondents settle on a unique answer to a number of items when the meanings of different answer options are mutually compatible. Accordingly, although scalar implicatures may render the exact interpretation of some questionnaire responses uncertain because of how they extend the impact of semantic context-sensitivity, their pervasive influence on the communicative dynamics of the questionnaire is mostly benign.

Compared to the significance of scalar implicatures, the potential effects of epistemic reflection that we have identified are minor because they concern only items with a ‘Don’t know’ answer option. Results from experimental epistemology indicate that when a subject is aware of a specific way in which her belief might be false despite her evidence, the subject is more likely to deny that her belief is knowledge. So, with respect to groups in which attention to such possibilities is highly common, there is a concern that respondents might opt for a ‘Don’t know’ answer despite having evidence supporting a different answer. Although this concern is mitigated by the nature of the question in two items with ‘Don’t know’ answer options, there remain two other items with ‘Don’t know’ answer options in which this concern is not mitigated by the question the item asks. Therefore, we believe that attention to epistemic error possibilities, pragmatic norms of communication, and semantic context-sensitivity, all belong in a description of factors shaping how respondents answer the SF-36.

References

- ALEXANDER, J., GONNERMAN, C. & WATERMAN, J. 2014. Saliency and Epistemic Egocentrism: An Empirical Study. In: BEEBE, J. (ed.) *Advances in Experimental Epistemology*. Continuum.
- ANDERSON, C., LAUBSCHER, S. & BURNS, R. 1996. Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke*, 27, 1812-6.
- BJORNER, J. B., DAMSGAARD, M. T., WATT, T. & GROENVOLD, M. 1998a. Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. *J Clin Epidemiol*, 51, 1001-11.
- BJORNER, J. B., THUNEDBORG, K., KRISTENSEN, T. S., MODVIG, J. & BECH, P. 1998b. The Danish SF-36 Health Survey: translation and preliminary validity studies. *J Clin Epidemiol*, 51, 991-9.
- BROWN, C. A., CHENG, E. M., HAYS, R. D., VASSAR, S. D. & VICKREY, B. G. 2009. SF-36 Includes Less Parkinson Disease (PD)-Targeted Content but Is More Responsive to Change than Two PD-Targeted Health-Related Quality of Life Measures. *Quality of Life Research*, 18, 1219-1237.
- BUCKWALTER, W. 2014. The Mystery of Stakes and Error in Ascriber Intuitions. In: BEEBE, J. (ed.) *Advances in Experimental Epistemology*. Bloomsbury Academic.
- CARSTON, R. 2002. *Thoughts and utterances*, Oxford, Blackwell.
- CECH, D. J. & MARTIN, S. T. 2012. Chapter 5 - Evaluation of Function, Activity, and Participation. In: CECH, D. J. & MARTIN, S. T. (eds.) *Functional Movement Development Across the Life Span (Third Edition)*. Saint Louis: W.B. Saunders.
- CLARK, H. H. & SCHOBBER, M. F. 1992. Asking questions and influencing answers. In: TANUR, J. M. (ed.) *Questions about Questions*. New York: Russel Sage.
- COLLINS, D. 2003. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res*, 12, 229-38.
- CONRAD, F. G. & BLAIR, J. 2004. Data Quality in Cognitive Interviews: The Case of Verbal Reports. *Methods for Testing and Evaluating Survey Questionnaires*.
- GRICE, H. 1989. *Studies in the Way of Words*, Harvard University Press.
- GUILFOYLE, M. R., SEELEY, H. & LAING, R. J. 2009. The Short Form 36 health survey in spine disease--validation against condition-specific measures. *Br J Neurosurg*, 23, 401-5.
- HANSEN, M. B., PETERSEN, E. N., POULSEN, A. & SALÈS-WUILLEMIN, E. 2017. When Belief Ascriptions Are About More Than What Is on Someone Else's Mind. *Discourse Processes*, 54, 655-669.
- HAYS, R. D. & SHAPIRO, M. F. 1992. An overview of generic health-related quality of life measures for HIV research. *Qual Life Res*, 1, 91-7.
- HORN, L. 1972. *On the Semantic Properties of Logical Operators in English*. PhD Dissertation, UCLA.
- HORN, L. 2004. Implicature. In: HORN, L. & WARD, G. (eds.) *The Handbook of Pragmatics*. Oxford: Blackwell.
- HORN, L. 2006. The border wars. In: VON HEUSINGER, K. & TURNER, K. (eds.) *Where Semantics Meets Pragmatics*. Oxford: Elsevier.
- HUANG, Y. 2017. Neo-Gricean Pragmatics. In: HUANG, Y. (ed.) *The Oxford Handbook of Pragmatics*. Oxford University Press.
- ICHIKAWA, J. & STEUP, M. 2018. *The analysis of Knowledge* [Online]. Available: <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/> [Accessed].
- ISRAEL, M. 2004. The Pragmatics of Polarity. In: HORN, L. & WARD, G. (eds.) *The Handbook of Pragmatics*. Blackwell.
- KENNEDY, C. 2001. Polar Opposition and the Ontology of 'Degrees'. *Linguistics and philosophy*, 24, 33-70.
- KENNEDY, C. & MCNALLY, C. L. J. L. 2005. Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. 81, 345 - 381.

- KENNEDY, S. 1999. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison (1st ed.)*, Routledge. .
- KROSNICK, J. & PRESSER, S. 2010. Question and Questionnaire design. In: MARSDEN, P. & WRIGHT, J. (eds.) *Handbook of Survey Research*. Second ed.: Bingley: Emerald group publishing limited.
- LEVINSON, S. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*, Cambridge, Mass.: MIT Press.
- LEWIS, D. 1996. Elusive knowledge. *Australasian Journal of Philosophy*, 74, 549-567.
- MALLINSON, S. 2002. Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Soc Sci Med*, 54, 11-21.
- MCHORNEY, C. A., WARE, J. E. & RACZEK, A. E. 1993. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*, 31, 247-263.
- NAGEL, J., SAN JUAN, V. & MAR, R. A. 2013. Lay denial of knowledge for justified true beliefs. *Cognition*, 129, 652-661.
- PEPPER, S. C. 1981. Problems in the quantification of frequency expressions. In: FISKE, D. W. (ed.) *Problems with language imprecision, New directions for Methodology of Social and Behavioral Science*. San Francisco: Jossey-Bass.
- RECANATI, F. 2010. *Truth-Conditional Pragmatics*, Oxford, Oxford University Press.
- SCHWARZ, N. 1995. What Respondents Learn from Questionnaires: The Survey Interview and the Logic of Conversation. *International Statistical Review*, 63, 153.
- SCHWARZ, N. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*, Psychology Press.
- SCHWARZ, N. & HIPPLER, H. 1991. Response alternatives: the impact of their choice and ordering. In: BIEMER, P., GROVES, R., MATHIOWETZ, N. & SUDMAN, S. (eds.) *Measurement error in surveys*. Chichester: Wiley.
- SCOTT, K. M., TOBIAS, M. I., SARFATI, D. & HASLETT, S. J. 1999. SF-36 health survey reliability, validity and norms for New Zealand. *Aust N Z J Public Health*, 23, 401-6.
- SEUREN, P. A. M. The structure and selection of positive and negative gradable adjectives. 1978.
- STANLEY, J. 2000. Context and logical form. *Linguistics and Philosophy*, 23, 33.
- WARE, J. E., JR., PHILLIPS, J., YODY, B. B. & ADAMCZYK, J. 1996. Assessment tools: functional health status and patient satisfaction. *Am J Med Qual*, 11, S50-3.
- WARE, J. E., JR. & SHERBOURNE, C. D. 1992. The MOS 36-item short-form health survey (SF-36). Conceptual framework and item selection. *Med Care*, 30, 473-83.
- WELDRING, T. & SMITH, S. M. S. 2013. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health services insights*, 6, 61-68.
- WRIGHT, D., GASKELL, G. & O'MUIRCHARTAIGH, C. 1994. How much is 'quite a bit'? Mapping absolute values onto vague quantifiers. *Applied Cognitive Psychology*, 8, 17.
- WÄNKE, M. 2002. Conversational norms and the interpretation of vague quantifiers. *Applied Cognitive Psychology*, 16, 301-307.
- AARONSON, N. K., ACQUADRO, C., ALONSO, J., APOLONE, G., BUCQUET, D., BULLINGER, M., BUNGAY, K., FUKUHARA, S., GANDEK, B., KELLER, S., RAZAVI, D., SANSON-FISHER, R., SULLIVAN, M., WOOD-DAUPHINEE, S., WAGNER, A. & J. E. WARE, JR. 1992. International Quality of Life Assessment (IQOLA) Project. *Quality of Life Research*, 1, 349-351.