Volume Title: Explorations in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90337-0
ISBN13: 978-0-226-90337-8

Volume URL: http://www.nber.org/books/wise09-2

Conference Date: May 2009

Publication Date: March 2011

Chapter Title: Comment on "Work Disability, Work, and Justification Bias in Europe and the United States"

Chapter Authors: Angus Deaton

Chapter URL: http://www.nber.org/chapters/c11947

Chapter pages in book: (312 - 314)

Van Soest, A., L. Delaney, C. Harmon, A. Kapteyn, and J. P. Smith. 2007. Validating the use of vignettes for subjective threshold scales. RAND, Labor and Population Working Paper. Available at: http://www.rand.org/pubs/working_papers/WR501/.

Comment Angus Deaton

I organize my comments around two facts or sets of facts that are at the core of this chapter:

- Fact 1: In matching surveys in Europe and the United States, people are asked to rate their own disability into five categories. The distribution of reports over those categories is very similar in Europe and the United States.
- Fact 2: In matching surveys in Europe and the United States, people are asked to rate other people's disability (using vignettes) into five categories. The distribution of reports over those categories is quite different in the two places; in particular, people in the United States are "tougher" in that they are less willing to admit that some conditions are disabling.

Kapteyn, Smith, and van Soest (henceforward KSvS) argue that fact 2 shows that Americans and Europeans use different scales to rank all disabilities, whether others or their own, a behavior that is known in the literature as "differential item functioning," or DIF for short. Once DIF is recognized, and the vignettes used to reinterpret the original responses in fact 1, we come to the conclusion that, in fact, the distribution of disability is worse in the United States than in Europe. I call this interpretation 1:

- Interpretation 1: Disability is worse in the United States than in Europe, but Americans have tougher standards, and so report the same levels as Europeans.

Fact 1 is misleading because Americans are tougher on themselves than are Europeans. The chapter makes a further contribution by allowing the extent of DIF to depend on whether or not someone is out of work—people might be less tough on themselves if they are not working, in effect a "justification bias." Their estimates suggest that this effect is present in the United States, but not in Europe. They conclude that "Americans use health as a justification for not working, whereas Europeans do not feel the need to do so."

Angus Deaton is the Dwight D. Eisenhower Professor of International Affairs and professor of economics and international affairs at the Woodrow Wilson School of Public and International Affairs and the Economics Department at Princeton University, and a research associate of the National Bureau of Economic Research.

As the authors note, this use of vignettes rests on two assumptions: "response consistency," which means you rank other people's conditions as you would your own, and "vignette equivalence," that respondents are equally "tough" or "lax" over all the vignettes that they are asked to rank. It is not obvious that either of these assumptions is correct, although the chapter lists two studies that provide some support for response consistency. The assumption is in principle testable; for example, by asking people to rate other people who have exactly the same conditions that they have themselves.

What we are talking about here is the extent to which people can feel each other's pain, whether they do it well, whether they want to do it, and whether they feel it as if it were their own. It is clear that people do not always care to do so, as demonstrated by my Scottish schoolmasters who, when administering corporal punishment to their pupils, and in visible pleasurable anticipation of the experience, would utter the ritual words, "This is going to hurt me more than it is going to hurt you." History is littered with more extreme examples. I know from my own experience that I am much more sympathetic to joint pain in others than I was before I had my own hip replaced, which would rule out vignette equivalence. I also know from my experience before the surgery that other people, even those who were trying hard, had little appreciation for the extent of my disability. There is also good scientific evidence that people have great difficulty in recalling even their own feelings or level of disability about previous conditions that have now been reversed; see in particular Smith et al. (2006) on the misremembering of colostomies. It is not clear that these examples are exactly relevant to the way that vignettes are used here, but they should certainly give us cause for concern.

We can think of response consistency as depending on the degree of empathy, whether respondents rank *other people's* conditions in the same way that they would rank their own if they were to have those conditions. For the vignettes to work for international comparisons, we require that there be no international variation in empathy, that Europeans are just as good or bad as empathizing with others as are Americans. There is a literature on whether women are more empathetic than men (Hoffman 1977; Eisenberg and Lennon 1983), but I know of know of no findings on international differences. However, my main point is not the factual one, but the logic of the vignettes. The use of vignettes *rejects* the assumption that people's self-reports of disabilities are internationally comparable, and *replaces* it with an assumption that their capacity for empathy is internationally comparable. Since the two assumptions are very similar, and similarly plausible or implausible, I do not see that anything is gained by replacing one by the other. The validity of vignettes depends on an assumption that is much the same as the assumption that they are designed to replace and I see no basis for accepting one and rejecting the other.

To see how this argument applies to the facts with which I began, I would propose an alternative interpretation of facts 1 and 2.

- Interpretation 2: The self-reports of own disability are accurate in the first place, and the distribution of disabilities in the United States and Europe is the same. The ratings of the vignettes are different because Americans are tougher on other people, without being tougher on themselves.

It seems to me that interpretation 1 and interpretation 2 are equally plausible—they provide contradictory views of the evidence, and I do not know how to choose between them. The correlation that KSvS report between differential responses to vignettes and national institutions for dealing with disability could just as well be due to differences in empathy as to differences in perceptions of own disability. Perhaps we need to add vignettes for empathy to the surveys, but then we would need vignettes for those vignettes, and so on ad infinitum.

For reasons already explained, I think that there must also be doubts about the assumption of vignette equivalence, that people are not differentially tough over different vignettes; for example, being more sympathetic to those where they have personal experience, or because different cultures recognize different disabilities. (Depression is a permanent condition in Scotland, but it is applied to the weather, not to people.) Vignette equivalence is also especially hostage to difficulties of translation; the problem is well-illustrated in the happiness literature by the fact that the English "happy," and the French "heureux," although exact translations, have different meanings and patterns of use in the two countries (see Wierzbicka 2004). It would seem relatively easy to relax vignette equivalence, at least in part, because there are always more vignettes than are needed to identify the model. Indeed, the model that is estimated in the chapter is clearly overidentified as written, and it would be a useful exercise to try to relax and test some of the assumptions. What I do not know is whether, if my main objections are met, identification will remain.

References

Eisenberg, N., and R. Lennon. 1983. Sex differences in empathy and related capacities. *Psychological Bulletin* 94 (1): 100–31.

Hoffman, M. L. 1977. Sex differences in empathy and related behaviors. *Psychological Bulletin* 84 (4): 12–22.

Smith, D. M., R. L. Sherriff, L. J. Damschroder, G. Loewenstein, and P. A. Ubel. 2006. Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology* 25 (6): 688–95.

Wierzbicka, A. 2004. Happiness in cross-linguistic and cross-cultural perspectives. *Daedulus* Spring: 34–43.