

Measuring Moral Judgment

Joshua May

Originally published on the [Experimental Philosophy blog](#) (05/26/2013).

I've been thinking a great deal lately about how best to study moral judgment. Let me say a few things myself, but I hope others will be able to chime in and share ideas, especially since I'm in the process of designing some studies.

What's best to measure?

So far it seems most researchers focus on measuring something other than what we might call purely "evaluative judgments," such as whether someone did something good or bad. This seems right to me since these don't necessarily constitute judgments about whether an action is right or wrong, which is paradigmatically a moral verdict. Moreover, this just seems to be the main target of most people doing research on moral judgment. For example, we want to know whether people think killing one to save five is morally wrong, not (merely) whether they think it involves doing some good. One might provide different answers to the corresponding questions.

So I think rightly the focus has been instead on what we might call "deontic judgments" such as whether an act is right/wrong, permissible/impermissible, etc. However, this then leads to issues about how to measure such judgments.

How best to measure it?

Some researchers present participants with a forced, dichotomous choice. They ask subjects to answer "Yes" or "No" to whether what the protagonist did was, say, permissible. ([John Mikhail's work](#) is a key example.) This has the advantage of quite clearly and straightforwardly yielding results that measure deontic judgments and places participants clearly into expressing one judgment or the other.

One potential problem with this, however, is that the forced-choice situation provides no option for participants to register uncertainty. (In fact, I suspect this has caused some trouble with some of this work. Maybe more on this some other time.) Another issue is that this is a "nominal" or "categorical" variable, which prevents a more fine-grained look at the data and makes certain statistical tests inappropriate.

Another approach is to present subjects with scales. Some researchers use something like:

0 = perfectly OK ... 9 = extremely wrong

(Compare [Haidt](#) and his various collaborators.)

One issue with this approach is that it's unclear, to me at least, whether wrongness really comes in degrees. Tom Hurka has recently argued that it does, at least in a certain sense (in a [blog post over at PEA Soup](#)). I have my doubts, which I've recently, albeit briefly, [expressed in a paper](#). I worry that the concepts of right and wrong (as opposed to good and bad e.g.) come in degrees in only a loose sense, as when we might say: "Murder is extremely illegal." In the comments section of his post, Hurka acknowledges that we might have to add "more seriously" to "wrong" to express a concept that comes in degrees. I wonder what other think, but I at least want to flag that this is an issue.

One could switch to a purely evaluative scale, such as:

1 = Very Good ... 7 = Very Bad

But, as I suggested before, I doubt this will shed much light on judgments about what is right and wrong, which I suspect are the primary targets of most researchers in this area.

Another Tack?

This all makes me wonder about the merits of taking a different approach. In other areas of experimental philosophy, researchers have tended to stick closer to the Likert-style scales, which are framed in terms of degree of agreement or disagreement, such as:

1 = strongly disagree, 2 = mildly disagree ... 7 = strongly agree

(Compare Liao et al 2012, ["Putting the Trolley in Order"](#))

Participants are then presented with statements and asked for their degree of agreement or disagreement with it (e.g. "It was morally wrong of Sam to steal the necklace.")

A similar approach has been pursued recently, albeit briefly, by Aaron Zimmerman in his [forthcoming commentary on Mikhail's book](#) in which he reports some new results. Zimmerman (along with John Caravello) presented participants with a scale that measures something like confidence in a claim (which is close to degree of agreement):

Should Jamie throw the switch?

(definitely yes) 1 - 2 - 3 - 4 - 5 - 6 (definitely no)

Focusing on something like agreement or confidence has the advantage of more clearly coming in degrees. So one needn't take a stance on whether deontic properties themselves are gradable.

Some worry that the usual mid-point of "Neither agree nor disagree" in Likert scales precludes equal intervals, since it seems to be a separate category. But some have taken to using "in between" for the midpoints, which seems a reasonable approach. And Zimmerman opted for no midpoint at all.

(I think there's another advantage of focusing on something like agreement, but that involves a further topic. I might take that up another time.)

These are just some thoughts I figured I'd make public. I'm curious to hear what others think about how best to study moral judgment. In particular, are there significant costs to measuring *degree of agreement* with moral claims rather than judgments about (alleged) *degrees of morality*?