Expectations and morality: A dilemma

doi:10.1017/S0140525X10001822

Eric Mandelbaum and David Ripley

*Faculty of Philosophy, The Future of Humanity Institute, University of Oxford, Oxford, OX1 1PT, United Kingdom; †Institut Jean Nicod, DEC-ENS, 75005 Paris, France.
Erich.Mandelbaum@philosophy.ox.ac.uk davewripley@gmail.com
http://www.fhi.ox.ac.uk/research/eric_mandelbaum
http://sites.google.com/site/davewripley

Abstract: We propose Knobe’s explanation of his cases encounters a dilemma: Either his explanation works and, counterintuitively, morality is not at the heart of these effects; or morality is at the heart of the effects and Knobe’s explanation does not succeed. This dilemma is then used to temper the use of the Knobe paradigm for discovering moral norms.

Knobe presents two kinds of theories that compete with his own: motivational bias theories and conversational pragmatic theories. He presents his own theory as a competence account. While we agree with his criticisms of the other accounts, we think his taxonomy is incomplete. We would like to suggest a different form of competence account, one that does not take morality as such to play a crucial role in these effects. (In this regard, we agree with Pheel and Sarkissian [2006], Macherly [2005], and indeed even Knobe and Mendlow [2004].) On our account, the effects of morality are a piece of a larger puzzle: Morality affects judgments of intentionality and related concepts only in virtue of its effects on expectations. Consequently, we think that anything affecting expectations will produce effects similar to those produced by moral norms. In fact, Knobe’s own account points to a similar conclusion, although he doesn’t acknowledge this.

According to Knobe’s competence theory, people’s moral norms influence their default expectations of others’ intentions, beliefs, values, causal roles, and so on, and these default expectations in turn affect participants’ judgments. Thus, he concludes, morality plays a deep role in explaining judgments in these various domains. But in this explanation, expectations are doing all the work: moral expectations have their effects only because they are expectations, not because they are moral. Thus, if Knobe’s theory is right, we should find effects similar to the effects cited here in cases that have nothing to do with morality, but instead involve participants’ non-moral expectations in parallel ways. And if this is right, it suggests that there is nothing specifically moral going on in the cases Knobe cites. These effects are, rather, effects of expectation, and expectations can be affected by both moral and non-moral factors (e.g., we expect people to have con-attitudes towards losing a game, although losing a game is not, normally at least, moral in any way).

Consequently, we think Knobe encounters a dilemma: Either his explanation of the effects he cites is correct, and then there is nothing especially moral at play here, but only an effect of expectations in general; or else his explanation of the effects is incorrect (in which case there may still be room for morality to play a distinctive role). Either way, Knobe finds himself in an awkward position; it doesn’t seem that his explanation of morality’s effects is compatible with the conclusion that moral considerations as such figure in our folk-psychological competence.

But we do not merely mean to present the dilemma. We take sides. We think Knobe’s explanation is substantially correct, and that the effects Knobe finds would follow from any expectations participants hold firmly enough, whether or not those expectations have a moral character. To see whether this is indeed the case, it is not enough to look at cases that involve moral factors. Similar cases involving non-moral norms must be constructed and tested.

As a step in this direction, we have conducted some preliminary studies involving variations on the CEO cases that involve non-moral norms. These studies were conducted using participants on Amazon.com’s Mechanical Turk website. We ran multiple studies attempting to measure possible non-moral effects on judgments of intention. In one such study, we used the following vignettes:

Normal case:

Two people are playing chess.

One of them considers moving her queen to square A6. She thinks, “If I move my queen to square A6, I will capture my opponent’s rook. But I don’t care at all if I capture my opponent’s rook; I think moving my queen to square A6 will allow me to checkmate in three moves.”

She moves her queen to square A6. Sure enough, she captures her opponent’s rook.

Did she “intentionally” capture her opponent’s rook?

Abnormal case:

The abnormal case was identical, except that the player allowed her own rook to be captured, instead of capturing her opponent’s rook. We expected that participants reading the abnormal case would be more likely to judge that the side-effect of the player’s move was intentional, when compared to participants reading the normal case. After all, it is normal to want to capture an opponent’s rook, and normal to want one’s own rook to remain uncaptured.

Although our results almost invariably trend in the expected direction, none actually reaches significance. (The closest result to significance arose from the vignettes given above; here, $\chi^2(11, N = 124) = 3.03, p = .08$.) For comparison, we also reproduced the original CEO cases using Mechanical Turk participants. Here, the results were highly significant: $\chi^2(1, N = 33) = 14.73, p < .001$. One possibility is that moral norms have a stronger effect on participants’ expectations than do non-moral norms (or at least the non-moral norms we tested). Another possibility is that Knobe’s explanation, which depends entirely on expectations, needs revision. Of course, either way, more systematic research is needed.

Our main point: One cannot only examine moral norms when judging whether Knobe’s data show an effect of morality. We must look at non-moral norms as well, to find just how broad the phenomenon is. In fact, Knobe has, in the past, thought similar things. In Knobe and Mendlow (2004), the authors propose that the kind of badness that affects intentional action judgments extends beyond just moral badness. They propose this in light of studies that seem to show similar effects involving clearly non-moral factors.

These theoretical possibilities matter for further work involving this effect. If indeed the effects Knobe finds are not specific to moral norms, then we must be careful not to interpret the effects as telling us about participants’ moral norms. For example, Inbar et al. (2009) use participants’ judgments of intentionality as a way to measure implicit moral norms. This is risky; although judgments of intentionality might tell us something about participants’ expectations in general, they cannot tell us which of those expectations are particularly moral and which are not. Use of intentionality judgments to measure implicit moral norms thus runs the risk of seeing moral norms where there are none.

NOTE

1. We thank Jesse Prinz, whose suggestion inspired these cases.

Norms, causes, and alternative possibilities

doi:10.1017/S0140525X10001834

Peter Menzies

Department of Philosophy, Faculty of Arts, Macquarie University, North Ryde, NSW 2109, Australia.

Peter.Menzies@mq.edu.au
http://www.phil.mq.edu.au/staff/menzies.htm
Abstract: I agree with Knobe’s claim in his “Person as Scientist, Person as Moralist” article that moral considerations are integral to the workings of people’s competence in making causal judgments. However, I disagree with the particular explanation he gives of the way in which moral considerations influence causal judgments. I critically scrutinize his explanation and outline a better one.

Knobe’s general explanation of the way in which moral considerations influence intuitive judgments goes like this: In judging causation, doing/allowing, intentional action, and so on, people select alternative possibilities to compare with what actually happens and their selection of these possibilities is influenced by their moral judgments. How does this idea explain the data about people’s causal judgments? Unfortunately, Knobe offers only the briefest hint in his Note 5, which suggests that moral considerations affect people’s causal judgments by influencing which counterfactuals of the form “If event c had not occurred, event e would not have occurred” they regard as true. This suggested explanation doesn’t work, however, for his own example in which Professor Smith’s action rather than the administrative assistant’s is regarded as the cause of the problem. This difference is not reflected in any difference in the counterfactuals people regard as true, since it is true that there wouldn’t have been a problem if either Professor Smith or the administrative assistant hadn’t taken a pen.

Luckily, Hitchcock and Knobe (2009) provide the missing elements of the explanation. Hitchcock and Knobe appeal to the finding in the literature on counterfactual availability that people are very inclined to entertain counterfactual hypotheses about what would have happened if a normal event had occurred instead of an abnormal one; and, by contrast, they are much less inclined to entertain counterfactual hypotheses in which normal events are replaced by abnormal ones. So people are willing to entertain the counterfactual about what would have happened if Professor Smith hadn’t taken a pen because it “mutates” an abnormal event into a normal event. By contrast, people are less willing to entertain the corresponding counterfactual about the administrative assistant’s action because it does not involve the privileged kind of “mutation.” Finally, by positing that people’s willingness to make a causal judgment “c caused e” goes hand-in-hand with their willingness to entertain the counterfactual “If c had not occurred, e would not have occurred,” they explain why people are more inclined to regard Professor Smith as the cause of the problem.

I suspect this explanation cannot be right for two reasons. The first is that the explanation involves an un economical hypothesis about the capacities involved in causal cognition. The explanation implies that people have an underlying competence for understanding counterfactuals that is linked to their objective core of the causal concept (the “causal structure” in Hitchcock & Knobe 2009). This competence is exercised when people understand counterfactuals of all kinds, including the counterfactuals about Professor Smith and the administrative assistant. Sitting alongside this competence, the explanation implies, is a psychological tendency to entertain some counterfactuals as “available,” a tendency aligned to people’s propensity to select certain events as salient causes. This hypothesis strikes me as implausible because of its doubling up of capacities involved in causal cognition.

My second reason for suspecting that this explanation can’t be correct is that empirical evidence casts doubt on the assumption that people’s causal judgments depend on their counterfactual judgments. Mandel and Lohman (1996), Mandel (2003), and Byrne (2005) cite experimental data that show that people’s causal judgments “c caused e” are dissociated from their counterfactual judgments “If c had not occurred, e would not have occurred.” The former go with judgments about sufficient conditions and productive mechanisms, whereas the latter go with judgments about enabling conditions and preventative mechanisms.

There is another way of developing Knobe’s general idea that moral considerations influence people’s causal judgments by way of their selection of alternative possibilities. In their classic work, Hart and Honore (1985) argue that the concept of actual causation originates in the situation in which a human action intervenes in the normal course of events and makes a difference in the way these develop. “The notion, that a cause is essentially something which interferes with or intervenes in the course of events which would normally take place, is central to the common-sense concept of a cause” (Hart & Honore 1985, p. 29). They argue that our judgments about what constitutes the normal course of events are guided context-sensitively – sometimes by what usually happens, and sometimes by social, moral, and legal norms. Their account readily explains why we regard Professor Smith’s action rather than the administrative assistant’s as the cause of the problem: for his action makes a difference to what happens normally – that is, in conformity with the prevailing norms – in a way that the administrative assistant’s does not.

Hart and Honore’s account of the way our causal judgments are shaped by moral considerations is better than Hitchcock and Knobe’s for several reasons: (1) Hart and Honore’s account captures in a seamless fashion the idea that causes are different from mere breakers for their effects. In contrast, it isn’t clear how Hitchcock and Knobe’s account captures this idea. Is it through the link with counterfactuals or through the rules about counterfactual availability? (2) Hart and Honore’s account doesn’t tie people’s causal judgments so closely with their counterfactual judgments, which is a virtue given the empirical evidence dissociating them. If it makes a link with counterfactuals, it is with counterfactuals that are based not on the actual world but on “normalised” worlds that abstract from the abnormal features of the actual world (Menzies 2007). (3) Hart and Honore’s account provides a more uniform account of the contrastive structure of actual causation. Many philosophers have observed that causal judgments have an implicit contrastive structure: the causal judgment “c caused e” has the implicit counterfactual structure “c rather than c* caused e rather than e*.” People typically select as the contrast elements c* and e* events that would normally have occurred if the abnormal actual events c and e had not occurred (Menzies 2009). This follows straightforwardly from Hart and Honore’s account, which incorporates the contrastive structure into the semantic content of causal judgments. If Hitchcock and Knobe’s account is to explain the contrastive character of causal judgments, it must do so through appealing to pragmatic or non-semantic rules about counterfactual availability.

Neither moralists, nor scientists: We are counterfactually reasoning animals

doi:10.1017/S0140525X10001846

Bence Nanay

Department of Philosophy, University of Antwerp, 2000 Antwerp, Belgium; and University of Cambridge, Cambridge CB2 1RD, United Kingdom.

bence.nanay@ua.ac.be / http://webh01.ua.ac.be/~bence.nanay
bn206@cam.ac.uk

Abstract: We are neither scientists nor moralists. Our mental capacities (such as attributing intentionality) are neither akin to the scientist’s exact reasoning, nor are they “suffused through and through with moral considerations” (Knobe’s target article, sect. 2.2, last para.). They are more similar to all those simple capacities that humans and animals are equally capable of, but with enhanced sensitivity to counterfactual situations: of what could have been.

Knobe presents us with a false dilemma on the level of the metaphors he uses: maybe we are neither scientists nor moralists. But he also presents us with a false dilemma when it comes to the two explanatory schemes he considers: The first one is that the