Unintentionally biasing the data: Reply to Knobe

ROBLIN R. MEEKS

Knobe (2003) wants to help adjudicate the philosophical debate concerning whether and under what conditions we normally judge that some side effect *x* was brought about intentionally. His proposal for doing so is perhaps an obvious one—simply elicit the intuitions of "The Folk" directly on the matter and record the results. His findings were a bit less obvious, however. When Knobe presented New York parkgoers with scenarios including either good or bad side effects, they tended to judge that the bad side effect was brought about intentionally and that the good side effect was not. In light of these responses, Knobe concludes that

> [p]eople's judgments depend in a crucial way on what *x* happens to be. In particular, it makes a great deal of difference whether they think that *x* is something good or something bad. (2003: 191)

He further explains this conclusion in terms of an underlying normative asymmetry, for according to Knobe the data suggests that "people are considerably more willing to blame the agent for bad side effects than to praise the agent for good side effects" (2003: 193). Hence, people's judgment that a side effect was brought about intentionally apparently rests, at least in part, upon how blameworthy they find the agent responsible for it.

Knobe's appreciably straightforward approach to this question does not settle the matter, however. In fact, his two scenarios gloss over a subtle but important distinction we often take into account when considering bad side effects. More importantly, the larger lesson to be drawn from his research is not that normative considerations influence our intuitions as to whether some side effect was brought about intentionally or unintentionally. Rather, simply raising that question can itself affect our evaluation of the side effect in question as either something good or something bad. As a result, Knobe's experiments unintentionally bring about

a side effect of their own—they effectively bias subjects' responses toward judging the given side effects more negatively than they might have otherwise.[1]

Ordinary language can indeed provide insight into these difficult matters as Knobe suggests, but his two scenarios don't take into account a subtle but important distinction that can influence our judgments of blameworthiness. Intuitively, in assigning blame it matters whether the relevant bad side effect is *part* of the action or a *result* of it. In Knobe's first example, subjects were told simply that the new, more profitable program harmed the environment and that the CEO was made aware of that possibility. Did the CEO's policy cause environmental damage *in* seeking higher profits, or did it do so *by* pursuing that more profitable strategy? Taking the timber industry as an example, it seems more egregious to tear up the landscape (with massive trucks, say) in the pursuit of lumber as opposed to the landscape ultimately suffering because of the lingering effects of deforestation. Similarly, in the second case, it's unclear whether the soldiers were killed in their being moved to the top of Thompson Hill or were relocated and then suffered casualties. In each case, the former reading construes the side effect as part of the action, whereas the latter one indicates that the side effect resulted from the action. I submit that Knobe would obtain quite different data if this distinction were made clear in both his help and harm scenarios. My hypothesis is that subjects would assign a higher level of blame if told that the damage and casualties happened *in* pursuing profits and hills. After all, we can make a fairly strong statement about a person's character by saying that she would step on her

[1] Speaking of bias, it's worth nothing that Knobe (personal communication) canvassed two Manhattan parks for subjects—Washington Square Park near New York University and Tomkins Square Park in the East Village. Both tend to attract young and often very liberal visitors, and hence I worry that these subjects might not be altogether neutral with regard to his scenarios. My guess is that in either location it wouldn't be hard to find those who would be quick to think the worst of corporations, their CEO's, and military officers—especially when considering corporate destruction of the environment and combat casualties.

own mother to get to the top. We don't, however, have an equally pithy saying for those who get to the top and then step on their mothers.

We encounter deeper difficulties when considering the "help" condition of both scenarios. In mulling over my own reaction to them (I am one of The Folk after all), I was reminded of J. L. Austin's admonition regarding our evaluations of actions in general—namely, that there is "*no modification without aberration*" (1961: 137). By this he means that when told that some person *S* did some action *A* we may feel entitled to ask whether *S* did *A M*ly or not *M*ly, for just about any modifier *M*. But in most cases we simply can't do so without indicating that the action was performed under special circumstances. To take Austin's example, if I sit in my chair in the normal way, I can't say either that I sat intentionally or that I sat in it unintentionally. We can, of course, imagine occasions when I would say that I sat down unintentionally, but on such occasions I would most likely say so to ward off the unfriendly accusation that I did so intentionally. (Perhaps in sitting I crushed a flower you left for me on my chair as an apology.)

Accordingly, by asking whether the beneficial environmental side effect resulting from the CEO's action was brought about intentionally, the context becomes one of special circumstances. It's an interesting fact of language that asking whether someone did a normally praiseworthy action intentionally effectively calls the purity of that action into question, such as when we express reservations about a giver's motives by asking whether her gift was given intentionally or voluntarily. Likewise for actions that are neither good nor bad. Consider running a race. If I win, it sounds odd to ask either whether I did so intentionally or that I did so unintentionally, barring peculiar contexts. And asking me if I lost the race intentionally sounds like an accusation of foul play. We simply can't always ask whether some action was done

intentionally or unintentionally without changing the context in a negative way, even when the action in question seems initially positive. What praise the CEO or officer might have earned was replaced by the misgivings that accompany the questioning of someone's motives—questions that Knobe's subjects could not resolve with the information provided in each scenario.

This observation connects to a second of Austin's lessons relevant here—namely what he calls the "*importance of Negation and Opposites*" (1961: 139). Depending upon the circumstance, the opposite of doing something intentionally might not be doing it unintentionally but doing it accidentally, carelessly, mistakenly, inadvertently, absent-mindedly, etc. If in my haste to catch the bus to the park I literally step on your toes, I may offer by way of apology that I did so unintentionally (or more accurately that I did so *accidentally*). Still, that doesn't mean that my actions before or after the injury were done intentionally. The modifiers 'intentionally' and 'unintentionally' might very well be—as Austin puts it—"fish from very different kettles" (1961: 139). Knobe's request that subjects determine whether a side effect was brought about intentionally or unintentionally relies on the questionable assumption that it must be one or the other.

In light of the above, the asymmetry in the influence of beliefs about the praiseworthiness and blameworthiness of actions that appeared in Knobe's results can be explained in the following way. Our attitude toward a specific side effect is intimately connected with our intuitions as to whether it was brought about intentionally or unintentionally because simply raising that question *converts* the context to a normative one in different ways. Subjects failed to assign a high level of praise for good side effects because taking into account whether they were brought about intentionally or unintentionally makes them suspect. Asking whether a

praiseworthy or even neutral action was done intentionally or unintentionally tends to sully it, and even innocent actions like sitting become much more ominous when we ask whether they were done intentionally. The contexts the result from such questions are those in which we think that someone has committed a wrong and needs to be held accountable, that someone has been wronged and needs to be made whole.

In the end we might go so far as to wonder exactly how empirical research like Knobe's can illuminate what in this case seems to be a largely, if not wholly, conceptual question. If the goal is to investigate the various ways in which we ordinarily talk about actions and their side effects, one need not frequent parks to do so. Simply considering what *oneself* would say when—though perhaps a questionable experimental methodology—should suffice.

Princeton Writing Program
Princeton University
Princeton, New Jersey 08544, USA
rmeeks@princeton.edu

References

Austin, J. L. 1961. "A Plea for Excuses." In *Philosophical Papers*. Oxford: Clarendon Press: 123-152.

Knobe, J. 2003. "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63: 190-194.