

Where Does Blaming Come From?

A REPLY TO KNOBE

Lawrence M. Solan[†]

[*Alvy and Annie are seeing their therapists at the same time on a split screen*]

Alvy Singer's Therapist: How often do you sleep together?

Annie Hall's Therapist: Do you have sex often?

Alvy Singer: Hardly ever. Maybe three times a week.

Annie Hall: Constantly. I'd say three times a week.¹

In my earlier essay, "Cognitive Foundations of the Impulse to Blame," I argued that blaming comes cheaply for people since the elements of the scenarios that most easily trigger blame are commonly used in cognitive processes that have little to do with moral attribution: causation, the recognition of bad outcomes, and sensitivity to the minds of others.² I relied in part on an ingenious experiment designed and run by Joshua Knobe, which demonstrates that people are highly sensitive to the differences between good and bad outcomes in their conceptualizations of the world.³ In his response to my essay, Knobe uses much of the same material, including his own experiments, to argue that I've gotten it wrong.⁴ He argues that people are primarily involved in the business of moral attribution, and the other cognitive processes are the derivative ones. Thus, the quote from *Annie Hall*. In this brief reply, I agree with Knobe that my analysis does little

[†] Don Forchelli Professor of Law and Director, Center for the Study of Law, Language and Cognition, Brooklyn Law School.

¹ ANNIE HALL (United Artists 1977). Quoted passage available at <http://www.imdb.com/title/tt0075686/quotes>.

² Lawrence M. Solan, *Cognitive Foundations of the Impulse to Blame*, 68 BROOK. L. REV. 1003 (2003).

³ Joshua Knobe, *Intentional Action and Side-Effects in Ordinary Language*, 63 ANALYSIS 190, 193 (2003).

⁴ Joshua Knobe, *Cognitive Processes Shaped by the Impulse to Blame*, 71 BROOK. L. REV. 931 (2006).

to explain asymmetries between our blaming, on the one hand, and giving credit, on the other, even though they both engage more or less the same conceptual primitives. However, Knobe's perspective leaves questions unanswered as well.

My original argument is as follows: Blameworthiness occurs when an individual causes a bad result, and becomes amplified when the perpetrator should have known better. All elements of the prototypical blame scenario – causation, bad outcome, and state of mind – are computed frequently in daily life for a host of reasons having nothing to do with blame. For example, conceptual development in children depends crucially and robustly on their ability to hypothesize about the states of mind of those from whom they learn, a phenomenon called “theory of mind” in the developmental psychology literature.⁵ This makes moral attribution cognitively inexpensive. In fact, it is so inexpensive that we sometimes manipulate facts either to increase or decrease the likelihood and severity of the blame impulse. In particular, when various interpretations of the facts are available, we tend to focus on those facts that enable us to reach a conclusion that is consistent with a desired result.⁶

Knobe argues that I've drawn the wrong conclusion. His claim is that “blame has had such a pervasive influence on our cognitive capacities that, even when we are not specifically interested in questions of blame, we often end up using cognitive processes that arose chiefly because of their role in making blame attributions.”⁷ Later, Knobe emphasizes that he intends a weak version of the hypothesis that moral attribution is cognitively prior to its elements:

Note that we are not here entertaining the absurd hypothesis that people's whole capacity for detecting causal relations arose out of a need to make assessments of blame. The idea is simply that certain aspects of this capacity – a capacity that presumably arose chiefly out of a need for prediction and explanation – may also have been shaped by a concern with attributions of blame.⁸

⁵ See Solan, *supra* note 2, at 1020 nn. 66-68 (providing references to some of the underlying psychological literature).

⁶ For discussion of the affective aspects of blame, see Mark D. Alicke, *Culpable Control and the Psychology of Blame*, 126 *PSYCHOL. BULL.* 556, 566-68 (2000); Mark D. Alicke, *Culpable Causation*, 63 *J. PERSONALITY & SOC. PSYCHOL.* 368 (1992).

⁷ Knobe, *supra* note 4, at 932.

⁸ *Id.* at 936.

I have no problem with the weak version of Knobe's hypothesis, other than its evolutionary bent. The question we are addressing is about how our psychology works today. To the extent that my earlier essay speaks in such historical terms, it is fairly subject to the same caution. Here, I will comment briefly on where I think the evidence now stands.

Knobe relies on the results of an experiment that he has designed and run that is both rich in its content and elegant in its simplicity. In the first experiment, a lieutenant orders a sergeant to send his squad to the top of a hill for strategic reasons. One version of the story has the sergeant responding that obeying this order will put the soldiers in the line of fire and at risk of losing their lives.⁹ The other version has the sergeant thanking the lieutenant for taking the soldiers out of the line of fire.¹⁰ In both versions, the lieutenant replies that he is not interested in the safety of the soldiers, but rather in the greater strategic decision. Subjects hearing this story blame the lieutenant for intentionally putting the soldiers at risk in the first version, but do not give the lieutenant credit for intentionally saving them in the second version.¹¹

In my earlier article, I argued from a similar experiment reported by Knobe that people are sensitive to the difference between good and bad outcomes.¹² But as Knobe correctly points out, the experiment also shows that assignment of intent – one of the elements of the impulse to blame according to my account – itself depends upon whether the scenario is blameworthy in the first place.¹³ What other explanation can there be for the asymmetry between the two versions?

Knobe's experiment surely shows that the blame impulse cannot be *reduced* to its cognitive elements.¹⁴ For if it could be, we would expect there to be a praise impulse as well. The individual who indifferently causes lives to be saved should be credited to the same extent as the individual who is

⁹ Knobe, *supra* note 3, at 192.

¹⁰ *Id.* at 192-93.

¹¹ *Id.* at 193.

¹² Solan, *supra* note 2, at 1017-18 (citing Joshua Knobe, *International Action and Side Effects in Ordinary Language*, 63 ANALYSIS 190 (2003)).

¹³ Knobe, *supra* note 4, at 934-35.

¹⁴ Psychologists generally agree that concepts cannot typically be reduced to knowledge of defining features. Knobe's experiment shows that *blame* is a case in point. For discussion, see GREGORY L. MURPHY, *THE BIG BOOK OF CONCEPTS* (2002); for a contrary view, considering *causation* as one of a number of primitives into which meaning may be decomposed, see RAY JACKENDOFF, *FOUNDATIONS OF LANGUAGE: BRAIN, MEANING, GRAMMAR, EVOLUTION* (2002).

indifferent to causing death is blamed. Clearly, he is not. Thus, it appears to be the case that our decisions to attribute intent to an individual are influenced by whether the outcome the individual has achieved is good or bad.

So far then, I agree with Knobe, and credit him with an important and subtle contribution to the psychology of intentionality. Nonetheless, there is a difference between the conclusion that particular attributions of intent are influenced by the outcome achieved (good or bad), and the conclusion that our concept of moral attribution exists in some way as a prerequisite to our internalized understanding of what intent means. I believe that the results of Knobe's experiment support only the first of these conclusions. In fact, they illustrate the second order effects that I described in my original essay: Once a subject concludes that blame should be assigned based upon the fact that the blameworthy individual caused a bad outcome, affective considerations contribute more and more to the degree of causation¹⁵ and, thanks to Knobe's contribution to learning in this area, apparently to the level of intent as well. That is, as the facts permit, we conceptualize situations to keep them consistent with theories we have formed, including theories of blame based upon an individual's having caused the death of others, as in Knobe's study.

The same holds true for causation, as already noted. Knobe describes in his response a second study in which he demonstrates that we are more likely to attribute causation when harm results from an omission by a person with responsibility to have performed an act than from the same omission by a person with no such responsibility.¹⁶ Here, however, there really is a symmetry with corresponding scenarios that result in good outcomes. We are much more likely to credit someone in a position of responsibility for improving things by not taking a harmful action than to similarly credit a passerby with no such responsibility. Thus,

¹⁵ For discussion of how people's judgment of causation can be influenced by their prior decision to attribute responsibility, see my earlier essay, Solan, *supra* note 2, at 1014-19 (discussing Lawrence M. Solan & John M. Darley, *Causation, Contribution, and Legal Liability: An Empirical Study*, 64 LAW & CONTEMP. PROBS. 265 (2001)).

¹⁶ Knobe, *supra* note 4, at 936-37.

this study appears to say more about differences between direct and indirect causation than about moral attribution.¹⁷

Moreover, it would be difficult to believe that causation, which we have known since Pavlov to be understood by animals as well as humans, to be motivated by considerations of moral attribution.¹⁸ Yet, once again, our attribution in any particular situation may well be influenced by considerations including the prior assignment of blame.

In conclusion, Joshua Knobe has provided important evidence to the effect that attribution decisions differ depending on whether one is being accused of doing something bad or praised for doing something good. In doing so, he has focused attention on the fact that moral attribution is not reducible to the presence of the elements that trigger it. My earlier essay presents no reason why it should not be, making Knobe's contribution valuable. Yet Knobe's essay does not go so far as to demonstrate that the primitives of blaming are somehow derived from blaming itself. Our essays, perhaps happily enough, will by no means be the final word on these important issues.

¹⁷ Languages often differ in their handling of direct and indirect causation to a greater extent than does English. For discussion, see Solan and Darley, *supra* note 15, at 295-96.

¹⁸ For recent studies on the concept of causality in primates, see MICHAEL TOMASELLO & JOSEP CALL, *PRIMATE COGNITION* 92-94 (1997) (describing causality in the cognition of apes).