

Philosophy 2²3³: Intuitions and Philosophy
Fall 2009
Tuesdays and Thursdays, 1pm - 2:15pm
Library 209

Hamilton College
Russell Marcus
Office: 210 College Hill Road, Room 201
email: rmarcus1@hamilton.edu

Class 20 - Folk Psychology and Juror Impartiality
Thomas Nadelhoffer, "Bad Acts, Blameworthy Agents, and Intentional Actions"

I. *Mens rea* and ascriptions of intentions

Knobe's research on intentional ascriptions demonstrated some oddities in our intentional ascriptions for side effects.

Subjects are more likely to ascribe intentions to an agent when they want to blame the agent, and in cases in which the side-effect was negative, than in cases where subjects do not want to blame the agent, or in which the side-effect was positive.

Nadelhoffer, in work we have not read, showed that the moral goodness of a side-effect can also affect intentional ascriptions.

Our intentional ascriptions seem to have an affective trigger.

Nadelhoffer ascribes a normative conclusion to Knobe, that ascriptions of intentional action properly track praise and blame assessments.

Knobe argues that the point of intentional ascriptions is to assist in the ascriptions of praise and blame, rather than to analyze the proper definition of 'intentional action' or to track or determine the actual state of mind of an agent.

We wondered whether such normative conclusions were warranted.

Nadelhoffer raises similar questions, at the end of his article.

But, for the purposes of this article, only the descriptive claim that people's intentional ascriptions are sensitive to their own affect is required.

Nadelhoffer applies these results about intentional ascriptions to questions about jury deliberations, and reports new results.

The problem with folk ascriptions of intentionality, Nadelhoffer argues, is that jurors are required to assess *mens rea* (or, culpable state of mind) of agents in order to determine their guilt or innocence.

For some crimes, including the most heinous crimes, the law differentiates among acts depending on the state of mind of the agent.

If an agent acts knowingly or purposely in committing murder, say, the crime is taken to be more serious.

For our present purposes, the jury instructions for deliberate meditation [contain] three elements... (a) that the defendant committed an unlawful killing (i.e., a killing that was not an accident or was not committed in selfdefense), (b) that the killing was committed with malice (i.e., the defendant either had an intent to cause death or caused the death intentionally), and (c) that the killing was committed with deliberate premeditation (i.e., the defendant thought before he acted and decided to kill after deliberation (155).

Thus, the law requires that assessments of intentionality be prior to ascriptions of guilt or innocence. If Knobe and Nadelhoffer's work is correct, such assessments seem impossible.

Results about the sensitivity of ascriptions of intentionality to moral assessment of an agent would be parallel to those concerning hindsight bias, as Nadelhoffer explains in fn 7.

Hindsight bias is the well-established phenomenon that people tend to overstate the foreseeability of an

action after it happens.

For example, researchers asked college students to predict how the U.S. Senate would vote on the confirmation of Supreme Court nominee Clarence Thomas.

Prior to the senate vote, 58-percent of the participants predicted that he would be confirmed.

When students were polled again after Thomas was confirmed, 78-percent of the participants said that they thought, at the earlier time, that Thomas would be approved.

As a result of hindsight bias, events that have *already occurred*, which is the case in *all criminal trials* - are judged to have been more likely to occur than they would have been judged *before their occurrence* (165).

Clearly, hindsight bias is a problem for jurors.

Nadelhoffer's goal is to show that ascriptions of intentionality are similarly problematic.

II. The Smith case

To consider a particular instance of the problem, Nadelhoffer examines the actual English case *D.P.P. v Smith*.

Smith, a thief driving a car with stolen goods, was attempting to escape from a police officer who was clinging to the side of his car.

The policeman was killed when he fell off of the car into oncoming traffic.

Here we have a case of a side-effect, since the thief's primary intention, presumably, was not to kill the officer, but merely to get away.

Nadelhoffer presented subjects two different scenarios (in a between-subjects experiment).

The first scenario (C1) was a streamlined description of the Smith case.

The second scenario (C2) was a description of a similar case, except that the driver of the vehicle is not a thief, but is attempting to escape from a car-jacker, again clinging to the car.

We are clearly more likely to blame the thief in the first scenario than we are to blame the driver attempting to escape from a car-jacker in the second.

But, Nadelhoffer designed the scenario so that the intentions of the two agents seem to be identical.

In both cases, the driver wanted to get away from the person clinging to his car.

Nadelhoffer compared subjects' responses to questions about whether the driver knowingly or intentionally killed the person clinging to the car, and whether the driver was to blame for the death.

If Knobe's results were to translate to Nadelhoffer's cases, we would expect that our different assessments of blame would transfer to different assessments of intentionality.

Indeed, Nadelhoffer found just that.

If we compare the results of C1 and C2, we see that even though the cases are identical in terms of the cognitive and conative considerations of the thief and the driver, the participants in C1 were more likely to say that the thief *knowingly* brought about the officer's death (75 percent) than the participants in C2 were to say that the diver knowingly brought about the death of the car thief (51 percent) - a statistically significant difference... (157).

III. Alicke and the CCM

So far, we have been considering only descriptive questions: do ascriptions of intentionality increase in cases of blameworthiness.

The cases we have seen are problematic, since the scenarios contain cases of identical, or relevantly similar, mental states combined with significantly different ascriptions of mental states.

Jurors are generally expected to evaluate evidence impartially.

Nadelhoffer's results seem to show an entrenched partiality in juror's ascriptions.

If jurors are liable to intentional ascription bias, then defendants are likely to be treated unfairly.

The partiality that Nadelhoffer describes is predicted by Mark Alicke's culpable control model, or CCM, of blame attribution.

In three of Alicke's studies cited by Nadelhoffer, Alicke shows that subjects ascribe more control to agents in cases where blame is spontaneously assessed.

In the first case, a homeowner is assessed greater blame for drinking two beers and shooting an intruder when it turns out that the intruder is his daughter's boyfriend than when it turns out that the intruder is a thief.

In the second example, based on the Bernie Goetz subway vigilante case, subjects blamed a subway shooter more when the victims turned out to be star athletes than when they turned out to be gang members.

In the third case, a driver who got into an accident while speeding was blamed more when he was trying to hide a vial of cocaine than when was trying to hide an anniversary present.

To the extent that the observer believes that the action in question is immoral, she will be inclined to look for explanations of the action that favor ascriptions of blame while at the same time over-looking explanations that do not (153).

Alicke's work supports that of both Knobe and Nadelhoffer.

It also extends their work by showing that ascriptions of blame are spontaneous, unconscious affective reactions.

Not only are subject inconsistent in their ascriptions of intentionality, those ascriptions are based on unconscious processes that are inimical to the rational, impartial deliberations we desire from jurors.

This spontaneous presumption of blame can cause the juror to *selectively look for evidence that supports blame attribution* while at the same time causing her to *overlook factors that might otherwise mitigate or exculpate blame or guilt* (159).

IV. The ameliorative project

Weinberg, Nichols and Stich, in "Normativity and Epistemic Intuitions," consider, in addition to the different descriptive and normative projects in epistemology, both an evaluative project, in which we see whether our actual reasoning matches our epistemic standards, and an ameliorative project, in which we attempt to rectify any shortcoming.

In the last section of his article, Nadelhoffer urges that such an ameliorative project is needed.

He rejects Knobe's contention that the oddity of our intentional ascriptions can be eliminated once we accept that the purpose of making such ascriptions is to facilitate praise and blame rather than assess the mental states of agents.

If ascriptions of intentionality are biased by our moral assessments of actions, then they are falling short of our cognitive expectations.

Even the mere observation that such tendencies are biases entails the normative conclusion that we should try to avoid them.

Nadelhoffer's work in this last section is fairly thin and derivative, relying mainly on work by Wilson and Brekke that show that such biases are difficult to overcome.

While the first step to avoiding bias is to know that it exists, research shows that there are more steps, and that these steps are difficult to take.

Wilson and Brekke identify four steps.

First, they must be made aware of the unwanted mental processes in question. Second, they must be motivated to correct the error. Third,...they must be "aware of the direction and the magnitude of the bias" (Wilson and Brekke 1994, 118). Finally, they must have sufficient control over their mental processes to be able to correct for the biases in question (160).

Nadelhoffer cites research showing that simple attempts to overcome bias are met with serious obstacles at all steps.

We can not merely demand that judges ask jurors to avoid such bias.

Even when people know that they are biased, they can not avoid it.

Consider another case of hindsight bias.

Researchers asked two groups to estimate the probability of flood damage caused by blockage of a city-owned drawbridge.

A control group was told only the background information known to the city when it decided not to hire a bridge watcher.

An experimental group was given this information, plus the fact that a flood had actually occurred.

Both groups were told that the city was negligent if the foreseeable probability of flooding was greater than 10%.

76% of the control group concluded that the flood was so unlikely that no precautions were necessary.

57% of the experimental group concluded that the flood was so likely that failure to take precautions was legally negligent.

In these two groups, we just see the standard hindsight bias.

In addition, a third experimental group was told the outcome and explicitly instructed to avoid hindsight bias.

These instructions made no difference: 56% concluded the city was legally negligent.

The ameliorative project appears doomed.

We have returned to Foley's dilemma: what do we do when our second-order knowledge of our bias conflicts with our first-order confidence in our judgments?