Forthcoming in *The Philosophers' Magazine*

# What Is Experimental Philosophy?

Joshua Knobe
*Princeton University*

Since the earliest days of analytic philosophy, it has been a common practice to appeal to intuitions about particular cases. Typically, the philosopher presents a hypothetical situation and then makes a claim of the form: 'In this case, we would surely say....' This claim about people's intuitions then forms a part of an argument for some more general theory about the nature of our concepts or our use of language.

One puzzling aspect of this practice is that it so rarely makes use of standard empirical methods. Although philosophers quite frequently make claims about 'what people would ordinarily say,' they rarely back up those claims by actually *asking* people and looking for patterns in their responses. In recent years, however, a number of philosophers have tried to put claims about intuitions to the test, using experimental methods to figure out what people really think about particular hypothetical cases. At times, the results have been extremely surprising.

Here I discuss applications of this new methodology to three areas of philosophy — the philosophy of language, the theory of action, and the free will debate.

## Philosophy of Language

One of the most influential appeals to intuition in recent analytic philosophy has been Saul Kripke's story of Gödel and Schmidt:

> Suppose that Gödel was not in fact the author of [Gödel's] theorem. A man called 'Schmidt'… actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the view in question [the then-popular 'descriptivist' theory], when our ordinary man uses the name 'Gödel', he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description 'the man who discovered the incompleteness of arithmetic'.

Readers of this story almost universally agreed that the word 'Gödel' did not, in fact, refer to Schmidt. Any theory which declared Schmidt to be the referent of 'Gödel' was therefore assumed to be incorrect.

A question arises, however, as to whether *everyone* shares this intuition or whether it is only shared by the kinds of people who normally read Anglo-American philosophy. The philosophers Edouard Machery, Ron Mallon, Shaun Nichols and Stephen Stich recently ran an empirical study to address this question. All subjects were given the story of Gödel and Schmidt. But the design of the study included a surprising twist. Some subjects were Americans; others were residents of Hong Kong. As expected, the American subjects shared the intuitions of most analytic philosophers. But the Hong

Kong subjects showed a quite different pattern of responses. Among subjects in Hong Kong, the majority said that the word ‘Gödel’ did indeed refer to Schmidt.

This recent result — along with similar results from studies in ethics and epistemology — suggest that Asian people may not share many of the intuitions on which widely accepted philosophical theories have been based.

**Theory of Action**

People typically distinguish between behaviors that are performed *intentionally* (e.g., picking up a glass of wine) and those that are performed *unintentionally* (e.g., spilling the wine all over one’s shirt). It has proven quite difficult, however, to say precisely wherein this distinction consists. People often have clear intuitions about whether or not a particular behavior counts as ‘intentional,’ but it is often unclear which particular aspect of the behavior is responsible for these intuitions.

In a series of recent experiments, Joshua Knobe has shown that people’s intuitions can actually be affected by the *moral* qualities of the behavior itself. He constructed pairs of vignettes that were similar in almost all respects but differed in their moral significance. In certain cases, subjects were far more willing to say that a behavior was performed intentionally when that behavior was morally bad than when the behavior was morally good.

So, for example, consider the following vignette:

> The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.’
>
> The chairman of the board answered, ‘I don’t care at all about harming the environment. I just want to make as much profit as I can. Let’s start the new program.’
>
> They started the new program. Sure enough, the environment was harmed.

Confronted with this vignette, most people say that the chairman *intentionally* harmed the environment.

But suppose that we replace the word ‘harm’ with ‘help.’ The vignette then becomes:

> The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.’
>
> The chairman of the board answered, ‘I don’t care at all about helping the environment. I just want to make as much profit as I can. Let’s start the new program.’
>
> They started the new program. Sure enough, the environment was helped.

Confronted with this second vignette, very few subjects say that the chairman intentionally helped the environment.

Does this result indicate that moral considerations are actually playing a role in our concept of intentional action? Or does it only show that our judgments can sometimes be distorted by feelings of blame? A variety of competing theories have been proposed, but no real consensus has yet emerged. Clearly, further research is needed.

**Freedom of the Will**

If everything we do is in some sense determined by the laws of physics, so that an enormous computer would in principle be capable of predicting our every action, can we still be morally responsible for the decisions we make? Many philosophers have held that the intuitive answer to this question is *no*. Thus, Robert Kane writes:

> In my experience, most ordinary persons start out as natural incompatibilists. They believe there is some kind of conflict between freedom and determinism; and the idea that freedom and responsibility might be compatible with determinism looks to them at first like a 'quagmire of evasion' (William James) or 'a wretched subterfuge' (Immanuel Kant). Ordinary persons have to be talked out of this natural incompatibilism by the clever arguments of philosophers.

Clearly, this claim about people's intuitions is an empirical one, and it should be possible to test it using standard empirical methods.

In a series of recent experiments, the philosophers Eddy Nahmias, Thomas Nadelhoffer, Jason Turner and Steve Morris did just that. Subjects were given the following story:

> Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25th, 2150 A.D., twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 PM on January 26th, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 PM on January 26th, 2195.

Subjects were then asked whether Jeremy was morally blameworthy for robbing the bank: A full 83% said yes. This result calls into question the widespread view that people ordinarily regard determinism as incompatible with moral responsibility.

Similar results were obtained in a study conducted by Rob Woolfolk, John Doris and John Darley. The researchers constructed stories about people who wanted to perform an immoral action and then were placed in circumstances in which they couldn't help but perform that action. (For example, one story concerns a person who wants to kill someone and then is forced to kill that very person by evil terrorists who give him a

'compliance drug' that makes it impossible for him to resist their orders.) Although subjects in these experiments knew that the agent could not possibly have done anything other than what he did, most felt that he was morally responsible for his behavior. This result again seems to undermine the claim that people are 'natural incompatibilists.'

**Conclusion**

Experimental philosophy is a relatively new area of research, and much important work still remains to be done. But the field appears to be growing extremely rapidly. We will surely see a number of surprising results in the years to come.

**Works Discussed**

Knobe, J. (2003). Intentional Action and Side Effects in Ordinary Language. Analysis, 63, 190-193.

Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). Semantics, Cross-Cultural Style. Cognition, 92, B1-B12.

Nahmias, E., Nadelhoffer, T., Morris, S., & Turner, J. (2004). Surveying Free Will: Folk Intuitions about Free Will and Moral Responsibility. Unpublished manuscript. Florida State University.

Woolfolk, R., Doris, J., & Darley, J. (2004). Attribution and Alternate Possibilities: Identification and Situational Constraint as Factors in Moral Cognition. Unpublished manuscript. Princeton University.