

ANSWERING OLD QUESTIONS IN A NEW WAY

SHAUN NICHOLS ON
EXPERIMENTAL PHILOSOPHY

by Trine Antonsen

Experimental philosophy reveals increasingly more about the physical processes underlying human action and reasoning. In addition to elucidating philosophical problems, this research also raises questions about what significance empirical data should have for philosophy. In March, the Centre for the Study of Mind in Nature arranged the conference *Neuroethics and empirical moral psychology* addressing questions concerning what neuroscience can teach us about moral psychology and what normative conclusion we should draw from this new knowledge. Shaun Nichols contributed to this conference with the talk *Moral Responsibility and Determinism: Investigating Folk Intuitions*. *Filosofisk supplement* asked him about experimental philosophy, the role of intuitions in philosophy and how research on folk intuitions can shed light on the hoary problems of freedom, responsibility and determinism.

What is experimental philosophy?

The primary thing that distinguishes experimental philosophy from psychology and social science is the topic. Experimental philosophers are interested in questions that psychologists haven't looked at. We look at the more elementary questions about the intuitions people have that are philosophically interesting and that philosophers recruit in their arguments and explanations. There are a lot of questions about mind and intuitions that philosophers find interesting, which are not asked by psychologists, and that is where it started. It's not as if psychologists are ill equipped to explain or understand these things; it is just that it has not been their focus. But there are now a number of psychologists who are interested in these issues. Marc Hauser's group at Harvard is one clear example. One thing I

would stress is that it is important to try to figure out ways to make experiments that will be philosophically interesting. We don't just want to duplicate things that the psychologists are doing, it somehow has to connect to philosophy.

And how would you compare experimental philosophy and more traditional, so-called armchair philosophy?

People view experimental philosophy as deeply opposed to armchair philosophy, but I think that is misleading. I definitely think of the projects that people engage in experimental philosophy as growing out of the issues that we think about traditionally in philosophy. The deeper points of contact are that a lot of the kinds of things that philosophers worry about and that philosophers are interested in from the armchair, those are the issues we're interested in looking at in a more social scientific way. But these techniques that we use are supposed to be a supplement. It is just one more thing to add, to try to fill out the picture of what is philosophically interesting.

Do you think that experimental philosophy is just a trend or do you think about it as "the new way of doing philosophy"?

I doubt that it is a trend in the sense that it is going to go away. I think it will, for some time to come, be an important area that contributes to philosophy. It is not a replacement for traditional philosophical discourse, but it gives us resources to answer questions in new ways. There are several different things that people do in experimental philosophy now. One example is that people are interested in the diversity of intuitions. I certainly used to take for granted that the intuitions I have about the philosophical cases that I was brought up on are representative, and that everyone shares my intuitions about these cases. It turns out that this is probably not the case. Looking at diversity between cultures and individual differences within cultures is one of the major research areas in experimental philosophy. It looks like there are differences between cultures; the sample of the East-Asian students and the samples of the Western students we looked at showed diversity in intuitions about philosophical thought experiments. Even within cultures it looks like different people have different intuitions about the same kinds of cases, and that they vary in systematic ways. This is a surprising and sometimes troubling fact to try to deal with when you try to build up a theory based on intuitions.



There is also a deeper, more interesting project, which is to try to figure out the sources of our intuitions. Often, finding out the source of your beliefs is useful for deciding whether or not you want to retain that belief. Think about Freud and his picture of religion as a kind of model for this, not because this picture of religion was right, but because he purposed an interesting challenge: If the reason we believe in God is because we want to, if we find out it is just wishful thinking, then that discovery does not prove that God does not exist, but it undercuts the warrant you have for the belief. I think that a lot of experimental philosophers are suggesting the same things, only in different domains. This is how I view Joshua Greene's work on moral intuitions. What he is trying to do is to suggest that the moral intuitions we have about trolley cases don't have the kind of rational status that many philosophers thought they did, because they have spurious origins.¹ These kinds of cases seem to me to be really interesting projects in experimental philosophy.

How do you respond to scepticism about experimental and empirical philosophy coming from the naturalistic fallacy-problem, that you cannot derive moral oughts from empirical facts?

My initial reaction is that a lot of work in experimental philosophy doesn't try to go from "is" to "ought". A lot of it is purely descriptive, just trying to figure out what the folk theory is and what psychological mechanisms that generate particular intuitions. But I also think that you can use descriptive facts to inform your normative judgments about what you should do or whether you should change your beliefs. You learn something from an experiment and you take that bit of

knowledge and feed it into the overall process you use to decide what is the right thing to do. The experiment is not going to be a decisive, single fact that is going to change the way you think about what is the right thing to do. But knowing why you think a certain thing or why you think that certain actions are wrong might help you to decide what is the best overall theory of morality. So you don't get a quick move from "these are the facts" to "therefore this is what you should do", but knowing the facts can help inform a process of reflective equilibrium to get to the right thing to do. So that is how I would draw the connection between experimental work and the ethical work.

In your talk about determinism and responsibility at the conference, and also in several papers², you have presented data from experiments on human intuitions and responsibility. Can you, in short, explain your findings?

The first set of experiments that I did on the responsibility questions was with Joshua Knobe. We described two universes, Universe A and Universe B where A was deterministic and B indeterministic with respect to human choice. We tried to use natural language and things that were familiar to people, so we didn't use words like "determinism", but we tried to capture the idea of determinism in natural language. First of all we asked people which universe was most like ours. We found that most people, at least in our sample in the West, said that our universe is most like universe B, the indeterministic universe.

Next we asked different versions of questions about responsibility. One version of the question was a totally abstract version that basically said: "In universe A, the deterministic universe, are people fully morally responsible?" We found that people tended to say that people in universe A are not fully morally responsible. This suggests that at least at some level people think that determinism precludes moral responsibility. We then asked questions that were emotionally salient like: "Bob lives in a deterministic universe and he kills his wife and children. Is he fully morally responsible?" Here people were more likely to say that Bob was fully morally responsible.

In another experiment, we tried to control the concreteness of the cases. We looked at two cases: a serial tax cheater and a serial rapist. We asked whether the serial tax cheater in a deterministic universe was fully morally responsible and we asked the same question about the serial rapist. We found that in the tax cheater case, which was presumably low affect, people tended

• Shaun Nichols

to say that he was not responsible in a deterministic universe, but in the serial rapist case where the affect was really powerful, people tended to say that he was fully morally responsible. So it looks like affect plays a role in modulating responses to whether or not people are responsible in determinism.

Where do you think this diversity in response comes from? Would you call it an affective bias or an affective competence?

My current guess, although this is just based on one experiment, is that it is the *bias* that is driving the effect. The reason why I think this is that when you first ask people about indeterminist universes you know that people will say that they think that our universe is indeterministic with respect to choice. Then you ask them something about the tax cheater and the rapist, and pretty much everybody respond that both of them are fully morally responsible in indeterminist universes. Secondly, you move to the determinist universe and ask people the same questions and you get a split; people say that the tax cheater is not fully morally responsible, but the rapist is. So when you have decreased the affect, people say that the tax cheater is not morally responsible in the determinist universe, even though in the indeterminist universe people think he is responsible. So it looks like affect is somehow biasing or distorting our natural response. Now, this is just one experiment and there are other ways of explaining it, and I hasten to emphasize that even if it is bias that generates this effect, and even if it is an affective bias, it might turn out at the end of the day that we are better off with it and we should embrace this view that people are responsible under determinism.

What do you think would happen if determinism were revealed as true?

I think this really is a much more complicated question than people have recognized. My guess is that people would be more apt to give up the view that responsibility conflicts with determinism than they would be to give up the view that people were responsible. The idea that people are responsible would have more staying power. One bit of evidence that we have on this is from a study I did with Adina Roskies. In one condition we asked them to imagine a universe where scientists believed that determinism was true, though again, we didn't use the term "determinism" but described determinism in a non-technical way. When we asked whether people in that universe were responsible or not, our subjects replied that they thought the

latter was the case. Then we used exactly the same text except we didn't say "imagine a universe where scientists believe determinism is true", but "many scientists believe determinism is true", so that subjects would think that this was the case on earth, and we asked: "Assume these scientists are right, do you think people are morally responsible?" Here subjects tended to say that people would still be responsible. So when you make it real for people and anchor it in their own thinking about the world it seems like people will revert to a kind of compatibilism about responsibility.

Again, this is just one study, but I think there might be general reasons to expect that the idea that people are responsible for their actions is deeply ingrained in us and that idea would be harder to dislodge than the idea that determinism is incompatible with responsibility.

You and many other experimental philosophers put great emphasis on intuitions. What are intuitions?

All the people I have been working with in experimental philosophy, like Stephen Stich, Joshua Knobe and Jonathan M. Weinberg, think of intuitions as spontaneous judgments. They are the responses people give to questions on minimal reflection. Sometimes people characterize intuitions as non-inferential judgments, but for the purposes of experimental philosophy that is probably a needlessly restrictive view, because it might turn out that a lot of the intuitions are produced by tacit inference, inference that occurs below conscious access, and that is really critical to psychological explanations of how you arrive at the intuition. Where intuitions come from will vary. Presumably they sometimes come from things that do involve lots of reasoning. In other cases much tacit reasoning may not occur at all.

When people try to explain their intuitions they often say: "It is just how I feel" or "I can feel that it is right". How are intuitions related to emotions? Are they only emotional biases or can they also come from reasoning?

I think we can have emotions without thereby having intuitions. You can feel sad without having an intuition that something is wrong or right. I regard the notion of intuition that is connected to people's feelings the way you describe it, as a different notion than the notion that people in philosophy have been interested in, which is the spontaneous judgment that you make about a case. I do think it is important to distinguish between intuitions and emotions in this way because

Answering old questions in a new way

sometimes people say that to have an emotion is to have an intuition. If you think of intuitions as something that allows you to make a distinction between what is right and wrong, what is a case of knowledge and what isn't, you are doing more than just having a feeling. You're making a distinction that is more fine-grained than what you get with emotions.

So you are saying that having an emotion is not having an intuition, but is having an intuition having an emotion?

I don't think so. You can have intuitions that aren't emotions. The clearest case is probably when one gives somebody a thought experiment about knowledge and they say "yes, that's knowledge" or "that's not knowledge". That is not an emotion. That is something else.

What is the epistemic role of intuitions?

This is something I worry about because I want to be able to retain a lot of the intuitions I have about morality. So for example I have an intuition that it is wrong to push somebody off the tracks in front of the trolley and I want to be able to retain it. But the fact that I have the intuition all by itself is not sufficient to justify that. I can't go from "I have this intuition that it is wrong to push the guy" to "It is wrong to push the guy". I think a real difficult issue is how to figure out how we should, in moral philosophy, use our intuitions to build our moral theories about what is right and wrong. I'm hopeful that we can build theories that aren't entirely ad hoc and self-indulgent, that will allow us to keep things like "it is wrong to push the guy to save five people".

In cases such as incest or killing babies our intuitions tell us that it is wrong. If someone tells you that these intuitions are completely wrong and that you have to give up these intuitions, I claim that you wouldn't. My point is that we do trust our intuitions.

Yes, we do.

Do you think we should?

I think it is a perfectly reasonable strategy if you don't have anything else to trust than your intuitions. I think intuitions and spontaneous judgments often turn out to be right in all kinds of cases. But when we are doing philosophy we should critically examine these intuitions we have and critically examine our commitments.

At least we have to put on the table whether or not to trust the intuitions. But in everyday life, to be sceptical about intuitions generally seems to be a difficult way to live.

But again, sometimes we have conflicting intuitions, so how can we know which one to trust?

This is the thing I find most promising about experimental philosophy. Where we have conflicting intuitions, if we can see that one of the intuitions comes from a bad source, but we don't have any reasons to think the other one comes from a bad source, that gives us a key bit of information that helps us to make a decision about which intuition to retain.

How exactly can we figure out which intuitions have a bad source and which do not?

If you could find that somehow the process that leads you to having an intuition about a case is one that leads to lots of false beliefs in other kinds of cases that are similar on the same dimensions, then there is reason to be sceptical about the intuition.

Peter Singer is asking why we should regard intuitions as having any normative force if they are just biological residues of our evolutionary history. How do you respond to that?

I think this is a serious question and it deserves an answer, but I don't think it is going to be easy to answer. I think that in some cases, like aesthetics, if it turns out for me that the reason I like a particular kind of music is because of something about my evolutionary history, or because of my specific cultural upbringing, or something like that, that is not going to diminish my appreciation of the music I like, and I don't think it should.

I don't think these normative judgments are undercut because they have some natural explanation. Of course in ethics the issues are much more serious, but it is not obvious that you can't give your intuitions considerable weight in making these kinds of decisions. But I think there *is* a danger that by building up your theories based on your intuitions you are doing something that is kind of self-indulgent; you are allowing yourself to go on with what seems right to you and knowing that this has a particular kind of evolutionary source that does not seem to reflect the way things are now, which I think is threatening. I do worry about it,

• Shaun Nichols

but so far I don't worry enough so that I'm willing to give up the intuitions. I am still holding out some hope that there's still some kind of justification for them. But I regard this as a very serious problem.

*In your book *Sentimental Rules* you raise objections against moral sentimentalism and neosentimentalism. What distinguishes your theory from the sentimentalist account?*

I tend to think about neosentimentalism as an over-intellectualised view of moral judgment. So that view (or the most prominent version of it) is that to judge an action as wrong is to judge that it would be appropriate to feel guilty if you did the action and it would be appropriate to resent somebody else who did the action. The problem for neosentimentalists is that it seems that kids are making moral judgments very early. If you use the moral/conventional-task, they do it just after the age of three. They are not very good at judging the appropriateness of guilt and resentment, so it looks like this is reason to say that moral judgment is something that is more basic than the judgments of appropriateness of guilt and resentment. The account that I give does not rely on any of these rich understandings of the appropriateness of emotions being required for making moral judgments. So that is how my view differs from the neosentimentalist view.

The "old" sentimentalist view is really different, and interestingly different from the story that I tell. The classic view you get from people like Hutcheson is that to judge something as wrong is to have a feeling of disapproval. The story that I tell is that it is not enough to just have the feeling, you actually have to have a rule represented, and the feeling plays a role in somehow giving that rule special salience or importance. In Hutcheson's account it's just to have this sentiment of disapproval, but the sentiment of disapproval is never defined very precisely. This gives it a flexibility that has made it very difficult to refute because anytime you say that an agent is not having the sentiment but is still making a moral judgment, you can say that the agent is not really having the sentiment of disapproval. That is why I love James Blair's theory of moral judgment – it is like the Hutcheson theory, but he is very precise about the emotion. He says roughly that to have a moral judgment is to have the feeling caused by the VIM (violence inhibition mechanism), and that is triggered by seeing superficial cues of distress. So you can just counterexample that and show that it is not right. You can show that there will be instances where you see cues of distress, like a child crying after falling down and getting cut up, where you know there is no

moral judgment. You feel upset when you see it, and yet you don't make a moral judgment.

*In *Sentimental Rules* you claim that making a moral judgment needs two different psychological components, an affective mechanism and a normative theory, that are conspiring. What is this normative theory that must be part of our capacity in making moral judgments?*

I'm assuming it is a set of representations that we have to characterize the things that we regard as violations. It is a set of rules: don't do this, don't do that. It is a theory in a very loose sense; it is just a body of information that specify various actions as transgressions. I'm inclined to think that we shouldn't assume that it is innate. I think that we can explain a lot of the content of moral judgment, and the content of this normative theory, without assuming that it is innate. So the one element that seems to suggest the innateness account is that some of these rules seem to be pan-cultural. So the idea that you shouldn't rape, you shouldn't kill, you shouldn't steal; those things are found in every culture, so you might think that those rules must be innate. But there might be a cultural evolution story to be told that those kinds of actions are particularly likely to really upset people, and I think there are already reasons to think that norms that prohibit actions that are intrinsically likely to make people mad or be upsetting will be especially likely to have cultural force to catch on. And so that might be why it turns out that all cultures have prohibitions against killing, because if you have a prohibition against killing, that is going to stick around.

If moral judgment depends on both these two components, does this mean that making completely rational and non-affective moral judgments is not possible? Is the affective component necessary for making moral judgments?

Good question. Can people exhibit spontaneous normal moral judgments through purely rational processes? It looks like psychopaths are an empirical counterexample, because they, if you take the James Blair's line on this, have pretty much normal rational competence but deficient emotional competence, and this explains their deficient moral judgements. So it looks like rationality is not sufficient for moral judgment. But if you ask whether it is possible for a psychopath to talk about morality the way the rest of us do, surely you can teach them to do it. Probably lots of the psychopaths in these studies do figure out what you are supposed to say. The thing about these moral judgment tasks is that they are pretty simple and it would take three minutes to explain to the psychopath how

he should answer the questions so that he will “pass” the moral/conventional-task. If you phrase your question about having rational moral judgments rather as a question about whether it counts as a genuine moral judgment if it doesn’t have the emotions, then I don’t know how to answer it, because I don’t know how to decide independently what counts as a genuine moral judgment. If you say that what counts as a moral judgment is the stuff that most normal people do when

they say killing is wrong, then it looks like emotions are involved. If you take away the emotions you don’t get the same kind of thing. But if you say that what counts as a moral judgment is identifying certain actions and saying that these actions are wrong, and anyone who does that counts as someone making a moral judgment, then whatever the actions are, you can presumably teach somebody who hasn’t got the emotional systems to do that.

NOTER

1 See for instance Joshua Greene’s (2003) “From neural is to moral ought” in *Nature Neuroscience* 4.

2 Like, for instance, “Folk intuitions on free will” (2006) and “Moral responsibility and determinism: The cognitive science of Folk Intuitions (forthcoming), both available at < <http://dingo.sbs.arizona.edu/~snichols/>>

3 *Sentimental Rules* (2006). Oxford: Oxford University Press.

Vil du abonnere på **FILOSOFISK SUPPLEMENT?**

Filosofisk supplement kommer ut fire ganger årlig, og behandler alle slags emner som kan tenkes knyttet til filosofi.

Et årsabonnementet koster kr. 140,-

For å bestille send en mail med navn og adresse til:
filosofisk-supplement@ifikk.uio.no



Mer info finner du også på vår hjemmeside <http://foreninger.uio.no/filosofisk-supplement/>