

Free Will and Experimental Philosophy:

An Intervention

Tamler Sommers

Draft: Please do not cite without permission.

The following is a transcript of a conversation with an Experimental Philosopher Who Works on Free Will and Moral Responsibility (acronym: ETHAN). Three friends have asked Ethan to come for a brief “healing” session to get him to address some recurring problem with his work.

Friend 1: The first thing I want to make absolutely clear is that we’re here to help you. Everyone here is in your corner. We’re not here to be dismissive, to tell you that your projects are not “really philosophy” or that they have no value.

Friend 2: Not that you deserve all the fawning media attention you get either but...

Friend 3: Stop that. We’re here to help Ethan, not lash out.

Ethan: What exactly are all of you talking about?

Friend 1: We all got together and agreed that we needed to have a conversation with you. We think you’ve been in a...well... a self-destructive spiral recently. And we know you can pull yourself out of it, but first you have to recognize what’s happening. We don’t want things to get to a point where you’re handing out flyers in Washington Square park like a...

Friend 3: Isn’t that how this whole movement got started?

Friend 1: Ok, bad example. We don’t want him descending to where he’s camped out in the woods on the outskirts of some university just...

Friend 3: Again...

Friend 1: Fine! Let’s just get this over with, ok?

Ethan: Can someone tell me what the hell is going on?

Friend 3: Let's begin with the elephant in the room, the biggest problem of all, the one that's plagued your work from the beginning, the days of Nahmias, Nadelhoffer, Morris, and Turner. A lot of the other problems stem from this one. Now Ethan—what the whole justification for probing the “folks” about their intuitions? Why is that important?

Ethan: Philosophers appeal to intuitions all the time. That's how they defend their theories of free will. They appeal to intuitions in their arguments. I say that at the beginning of practically every paper I've done.

Friend 1: And rightly so, good. But you can't just probe for any intuitions you want, right? Which intuitions would be important to know about if that's your defense of the experimental approach?

Ethan: The intuitions about free will and responsibility obviously.

Friend 2: You just want their random thoughts about free will?

Ethan: You know what I meant. The intuitions at the center of the debate.

Friend 3: Right. And even more specifically, wouldn't you want to be testing for the same intuitions that the philosophers are appealing to in their theories?

Ethan: Of course, that's obvious.

Friend 2: If it's so obvious why aren't you doing it?

Ethan: What do you mean? I am doing it.

Friend 2: No you're not. Let me ask you something—how many, say, incompatibilist arguments are there that go like this?

1. If determinism is true, then intuitively there is no such thing as free will.
2. Therefore, determinism is incompatible with free will.

Does that sound like a good argument to you? Or would that be a just little question-begging?

Ethan: Well, that would be question-begging. Although Kane and Galen Strawson do say that people start out as naturalists incompatibilists and have to be talked out of it by the clever arguments of compatibilists.

Friend 2: We know, trust us. You quote the same damn passages in all your papers. You know where those passages come from? Introductions. They aren't part of any argument. They're throwaway lines. They play no role in incompatibilist arguments. If they weren't there, it would make no difference. Is that really what you want to test? Intuitions about introductory throwaway lines?

Ethan: Well, no but—I'm not sure I fully understand....

Friend 1: What he's trying to say is that you're testing intuitions on the wrong question. You're probing intuitions on the compatibility question itself. You're testing intuitions about the *conclusions* of compatibilist and incompatibilist arguments. But philosophers don't appeal to those intuitions. They appeal to intuitions about the *premises* of their arguments, about cases or narrower principles like transfer principles or PAP principles. We've made this point before...

Ethan: Ok, right, I see. You think that people's judgments about those principles or cases might be different than their judgments about the compatibility question.

Friend 3: Exactly. Now we're making progress.

Ethan: That's an interesting hypothesis, you should test that.

Friend 2: Here's an idea. After someone makes an objection to a study or a whole series of studies, how about you don't just deflect it by telling them they should test it? You know what psychologists never say when people raise objections to their studies? "That's an interesting hypothesis, you should test that." You want to know why they don't say that? Because they know that other people have their own lives, their own labs, their own careers, and they're not going to drop everything just because of the problems with another psychologist's experiment.

Ethan: But Josh Knobe says that all the time.

Friend 3: Look, Ethan. I know this is difficult. But Josh Knobe, important and influential as he is, much as we all admire him and love him, Josh Knobe is not God.

Ethan: I know that.

Friend 2: Say it then. Josh Knobe is not God.

Ethan: I just said it.

Friend 2: No you didn't.

Ethan: (fidgeting uncomfortably) Ok, fine. He's not.

Friend 2: Not what?

Ethan: (more fidgeting). Not God.

Friend 2: Who's not God.

Ethan: Josh Knobe, OK? Josh Knobe is not God. Satisfied?

Friend 1: The point is that this a glaring flaw with *your* experiments, *your* whole way of approaching this topic. *We* are not working on this issue right now and we have other commitments. And it's plaguing so much of your work these days. Look at this study that's just coming out by Florian Cova: "Judgments about Moral Responsibility and Determinism in Patients with Behavioural Variant of Frontotemporal Dementia: Still Compatibilists." What a cool idea, right? Testing patients with emotional deficits. We can learn so much—there are so many potential implications for Strawsonian approaches to moral responsibility to name just one. But what is he doing in the study? He's giving these patients the same questions that Nichols and Knobe gave in their original *Nous* article—questions that test the conclusions of arguments. And then he concludes that the patients are still compatibilists. But you can't say that they're still compatibilists unless you ask the right question. You're wasting so many golden opportunities.

Ethan: But those questions are industry standard now.

Friend 1 (to Friend 3): Avoidance.

Friend 2: Are you listening to us? That's our whole point: the industry standard is fucked. And anyway, Nichols and Knobe came out in 2007 for Christ sake. The seminal Nahmias et al paper

came out in 2006. How much of a standard can there be in a few years? You need to change the industry standard, or else every other future paper will have this exact same problem that undermines everything you're trying to do.

Ethan: What about the Chandra Sripada paper that just came out in *PPR*? He's testing for intuitions about manipulation cases. And then for intuitions about whether manipulated agents are acting from their 'deep self.'

Friend 1: That's exactly the kind of approach that you need to adopt—because it's testing *premises* of real philosophical arguments! But that paper is the exception. It strays from the industry standard and it should.

Friend 3: Let me point out something else. The original classic X-phi study—the famous Josh Knobe in Washington Square park experiment with the CEO and the environment—that worked precisely because it was testing the right intuitions. Intuitions about a particular case. No determinism, no ambiguous words. Did he intend to help or hurt the environment? That's all. There's a model of the kind of thing you need to be doing.

Friend 1: And this leads us to the second problem we think you need to face up to.

Ethan: What now?

Friend 1: These concepts that you're asking about in the studies are far too complex and ambiguous, they can be interpreted in so many different ways. So right now there's way too much noise in your experiments.

Ethan: Can you be just a little more specific?

Friend 3: Well, not to pick on Nichols and Knobe, but they can handle it. Let's go back to that study. They give their participants a description of a determined universe, we talked about the problems with that already. But then in the question, you ask: Can an agent, or can Bill, be "fully morally responsible" for their actions. Now "fully morally responsible," that's a strange term...

Friend 2: I've worked on this issue my whole career and I don't know what the hell that means.

Friend 1: The point is that fully morally responsible can be interpreted in a lot of different ways. I mean, setting aside the determinism question: is anyone ever fully morally responsible? External factors play *some* role in every action, right? Nobody disagrees with that.

Ethan: So how would you put it then?

Friend 3: That's a good question. To be honest, every time I teach this topic in an intro course it takes me at least 15-20 minutes to nail down the sense of moral responsibility that's at the center of the debate.

Ethan: What sense is that?

Friend 3: Moral desert, deserving blame or punishment independent of the consequences.

Ethan: But you can't say 'moral desert.'

Friend 1: That's right, you can't. That's a meaningless term to a 19 year old non-philosophy major. And even philosophers disagree about how to analyze moral desert.

Friend 3: In my class, we've already discussed non-consequentialism so I can use terms and concepts that wouldn't work for your surveys of non-philosophers. That's why I can get the concept across in 15 or 20 minutes. But that's certainly not the students' first association of the term moral responsibility if they have one at all. Look, it's a really tough concept—but as philosophers who work on this issue, *we* know what we mean for the most part, so we've forgotten all the ways that people interpret it.

Ethan: Ok, so what about 'blameworthy' and 'praiseworthy'?

Friend 1: Those are philosophers terms. People don't use that language outside of a class or a courtroom.

Friend 2: Nichols and Knobe is not even the worst example. How about Roskies and Nichols asking if people "should be morally blamed for their actions" in a determined universe. That's confusing on a bunch of levels.

Ethan: That paper was published in the *Journal of Philosophy*!

Friend 2: You're right, so it must be perfect. Can you stop avoiding the issue? It's very annoying. First of all, what do they mean by 'morally blamed'—have you ever heard anyone use that term in your life? Even worse, someone could think people should be blamed for consequentialist reasons even if they don't deserve it.

Friend 3: And this is our point. It's one thing not to get bogged down in the minutia of philosopher-speak. We're all in favor of that. We're not fastidious word-choosers, we don't get aroused by conceptual analysis. But it's another thing to play so fast and loose with terms that you have no idea what the subjects are thinking about when they answer. That's exactly the kind of noise you have minimize in a study.

Ethan: It's easy to just sit there in your armchair, criticizing our terminology and our work. But I haven't heard you suggest any better options.

Friend 2: Don't play the martyr with us. It's not our job to write your studies for you. The problem with you X-Phi people is that you too much unfair criticism, all these over-the-top hostile strawman objections by people who just want to trash your work because they feel threatened. So now you have a bunker mentality. You can't see that there might be legitimate serious problems out there.

Friend 3: Look, we understand that it's a tough problem, getting that language right. To be honest, I've thought about this a lot and I'm not sure how to phrase those questions. I think you need to get a little more clever and oblique. Maybe you can do some behavioral studies where someone gets punished at a cost for no other reason than they deserve it.

Friend 2: And keep it as close to real life as possible. For the love of God, no more planet Erta scenarios or hypothetical worlds where alien scientists have developed supercomputers to detect neuronal activity and predict everything that will happen.

Ethan: Philosophers use wacked out thought experiments all the time! You can't blame that on me.

Friend 2: Philosophers hold their big job meeting between Christmas and New Years in cold weather cities. Does that make it right? Moral responsibility is a *human* phenomenon—haven't you read Strawson for Christ's sake? Our intuitions respond to actual things that happen

to us, not to ludicrous science fiction fairy tales. We probably learn something from probing judgments about alien worlds and hypothetical fantasies, but nothing resembling what we want to learn.

Ethan: I actually agree with that. But look, this is a new field. We're learning on the fly. How do you think that we're going to make any progress if we don't try new things?

Friend 1: I'm so glad to hear you say that, Ethan. What it shows is that you're starting to recognize the flaws in your approach. You're making progress.

Ethan: Great, thanks. Now if you could just sound a little less patronizing...

Friend 2: You want less patronizing? I can handle that. I have a couple things to get off my chest and I promise you it won't sound patronizing.

Friend 3: Let's stick to the plan, OK? The next item on the agenda is how Ethan is starting to overreach a bit. Overstating the significance of certain results...

Friend 2: Overreach?? Are you freaking kidding me? I want to read something to you. You wonder why you have so many hostile unsympathetic critics, right? Well here's some prime ammunition for them. All of your problems distilled in a couple pages. This is from Knobe and Nichols in the new edition of the Kane free will anthology. They're discussing a possible objection: that different cultures might possibly have different intuitions about free will and responsibility than some college students at Florida State or wherever, and that certain intuitions in the debate may be shaped by cultural forces like the American emphasis on individual autonomy. You want to talk about patronizing? Here's how they respond:

We certainly agree that these are very plausible hypotheses [hey thanks!], but the empirical evidence thus far has not been kind to them. In a recent study, subjects from India, Hong Kong, Colombia and the United States were all presented with the abstract condition of the experiment described above (Sarkissian, Chatterjee, De Brigard, Knobe, Nichols & Sirker forthcoming). Strikingly, the majority of subjects in all four of these cultures said that no one could be fully morally responsible in a deterministic universe, and there were no significant cross-cultural differences in people's responses. Yet ordinary people, many of whom have never thought about these questions before, seem

somehow to immediately converge on one particular answer. In fact, we find this convergence even across four different cultures, with radically different religious and philosophical traditions. What could possibly explain this striking pattern of intuitions? [“Free Will and the Bounds of the Self.” Forthcoming. p. 4.]

Ethan: That’s an interesting result. What’s the problem?

Friend 2: Well, let’s break this down. “The empirical evidence has thus far not been kind” to the diversity hypothesis. What does this unkind empirical evidence amount to? One study. Is there a mass of empirical evidence that suggests otherwise? Yes. Is any of it discussed? No. But OK, that evidence doesn’t specifically concern the compatibility question (which is itself a Western obsession but never mind), so maybe they’re after something more specific. All right, let’s look at the study. They give the Nichols and Knobe deterministic scenario to university students in Hong Kong, India, and Columbia—university students, who supposedly represent the essence of every culture—then ask is it’s possible for someone to be morally responsible for their behavior in that deterministic universe. We already talked about the fundamental flaw in this approach--you’re testing intuitions about the conclusion of all the compatibilist and incompatibilist arguments. But set that aside for one second too. Pretend that’s not a problem. You’re basing this incredible conclusion—universal convergence on an ancient philosophical problem—on *one study that gets a null result!* Let me repeat that: the null result of a single study. That’s the evidence that “hasn’t been kind” to the diversity hypothesis. But it’s a null result—who the hell knows why you didn’t get significant differences? There are so many possibilities. Could something be wrong, I don’t know, with the methodology of the studies maybe? Could that be why you didn’t get significant differences? And we’ve talked about the problem with ‘fully morally responsible’ in English. What word are you using in Chinese? As far as I know, there is no Chinese word for moral responsibility in the sense that we’re talking about?

Ethan: The Hong Kong students were fluent in English. They did the study in English.

Friend 2: Even better. These Hong Kong students who are fluent in English are supposed represent ordinary people in a different culture with “with radically different religious and philosophical traditions.” And they’re supposed to interpret “fully morally responsible” just like

the Americans, and just like me I guess, even though I still don't have a clue what that means. And of course, all the participants in each country are (a) university students, (b) from high SES backgrounds, and (c) of university age. These are your "striking," "deeply puzzling and mysterious" results that cry out for explanation.. A null result from university students with same backgrounds from a study with ambiguous terminology that asks the wrong question. Do you see the problem? Do you see why people get mad at experimental philosophers?

Ethan: You seem to be taking this study personally. Do you have a dog in this fight or something?

Friend 1: I think he knows someone who wrote a book on the diversity of intuitions about moral responsibility.

Ethan: Who?

Friend 2: Nobody, that's not the point, let's stick to the issue.

Friend 3: We don't have problems with the study so much, more the wild claims that Knobe and Nichols are making about it. The study is, at best, an intriguing first step—it might even be suggestive if it weren't for the problem that it tests the wrong intuitions. But to conclude that "ordinary people, many of whom have never thought about these questions before, seem somehow to immediately converge on one particular answer" from that study is completely implausible.

Friend 2: Let's put it this way. You're always complaining about armchair philosophers who pull their claims about people's intuitions out of their you know what—and rightly so. But is what you're doing here any better?

Ethan: At least we're trying to test for intuitions in these other countries. That's something, isn't it?

Friend 2: It's something, but it might be worse in the end than what the armchair people are doing because you're bathing itself in the haloed pretense of real empirical support. But it's just a pretense in this case.

Ethan: (Breaks down, sobbing). You're right. Oh God I see it now! I'm so so soooooorry.

Friend 1: Good, good... let it out.

Friend 3: It's Ok, we're here for you.

Friend 2: This is bullshit, right?

Ethan: (stops sobbing immediately). Yes, I figured that's what I was supposed to do. But seriously, I get the point. Can we wrap this up?

Friend 1: I don't know, can we? What have we learned from this so far?

Ethan: Look, if you're waiting for me to really break down and start sobbing, that's not going to happen.

Friend 3: We don't expect that, we just want an acknowledgment that you understand that you have some problems and that you intend to address them. Again, we're your friends and you have our full support.

Ethan: Ok, yes, I'll try to address them.

Friend 2: Address what?

Ethan: All right, fine. First, I have to stop giving scenarios with deterministic worlds, and then asking if people can be morally responsible in them.

Friend 1: Because?

Ethan: Because when I do that, I'm not probing for the intuitions that philosophers are appealing to in their arguments. I have to start testing for intuitions about cases and principles, and investigating the source of those intuitions.

Friend 3: Exactly!

Ethan: Second, I need to nail down the sense of moral responsibility we're after. Not that you had any good ideas about the best way to do that but... that's my responsibility.

Friend 2: In a different sense of the word, yes it is.

Ethan: Ha ha, you should be headlining at Carnegie Hall. Third. I need to avoid outlandish thought experiments and bring my scenarios down to earth as much as possible. Fourth, I need to stop exaggerating the implications of the data from my studies, even though most social psychologists and cognitive scientists do that all the time.

Friend 1: Fair enough. But that doesn't make it right.

Ethan: Fifth and most important: the next time my friends invite me to a so-called "intervention" about my work, I need to tell them to get a life, and then go see a movie with my family. Is that an item on your agenda? It should be.

.