Why Insights in Evolutionary Moral Psychology Help Resolve Long-Standing Meta-Ethical Questions

12 November 2017

Normative statements are found in the earliest human literature, and they remain a central part of human discourse. Yet, pinning them down has proven remarkably difficult. When we say 'killing is wrong' or 'you ought to help others in need', we seem to be – as the label implies – comparing our behaviour to a set of norms. But what are these norms, and how do we know them?

For most of Western history, they were widely believed to come from God, as commands. In fact, the idea of behavioural norms and the idea of a powerful person to create and enforce them fit so seamlessly together that it's hard to imagine one without the other. And, of course, *revelation* is a straightforward explanation of how people came to know these norms.

Yet, as the progress of science has made it increasingly difficult to believe God exists, people continue to make and respond to normative statements. It seems that people are referring to something other than God's commands when making these statements, but what?

An answer might be that norms exist without God in an objective realm. This raises a problem, however. Since they are inaccessible to our senses, our only way of accessing them is through reason. But reason must start with facts, which are all we can access through our senses. (Excluding some form of transcendent intuition.)

David Hume dashed this method, though. Hume's objection is, somewhat unfortunately, I think, often labelled the *is-ought gap* and explained as 'one cannot derive an ought from an is'. This is a bit confusing; an 'ought' *is* a derivation. There's no inherent conflict between 'is' and 'ought'. When I say, 'he left half an hour ago, so he ought to be here any minute', I'm deriving an expected event from a known event; the 'ought' represents the derivation. What Hume is really pointing out is that we have no apparent method of deriving norms from facts. (Hence, it's more accurately described as the fact-norm gap.) We attempt it all the time in moral discourse, but it's not at all clear how it could possibly work.

We see this problem at work in Utilitarianism, arguably the most popular contemporary normative theory. Utilitarianism holds that one ought to maximise aggregate net happiness. (Or a slight variation hereof.) In other words, the norm to which all behaviour is compared to is happiness-maximisation. But how could we possibly come to know this?

Utilitarianism was initially developed by Jeremy Bentham out of an examination of human behaviour; humans act to maximise pleasure and minimise pain, Bentham found. John Stuart Mill later came to the belief that humans strive for happiness in a more complex fashion than pleasure-seeking. More recently, many Utilitarians have pointed to discoveries that some animals have sentience.

These are all factual claims, but there's no apparent justification for thinking they say anything about norms.

(In fact, Utilitarianism takes the observation that people seek *their own* happiness and infers that they ought to seek *aggregate* happiness, even at the expense of their own happiness. Not only does it make an unjustified inference from fact to norm, but it makes a subtle change, with important consequences. Consider a man who faces the option of sacrificing himself in order to increase aggregate happiness. Let's assume he's a Utilitarian, so he does it. But this contradicts the initial fact from which Utilitarian norms were derived, namely that people seek their own happiness. Which leads to the paradoxical situation that after a two-way inference – first from facts to norms, and then back to facts – we end up with the opposite fact of that which we started with.)

There are surely more sophisticated versions of Utilitarianism, but it's difficult to see how any such endeavour could possibly avoid Hume's fact-norm gap. In fact, we could take a knife and slice the theory of Utilitarianism into two parts, a factual part and a normative part. There's no apparent way to connect the two.

This seems to rule out the possibility of our normative statements referring to an objective set of norms; we would have no way of knowing them even if they did exist, so they can't be what we're referring to.

<div align="center">***</div>

In recent years, moral psychologists have attacked this question from a different angle. Instead of examining how humans behave and trying to figure out what that says about norms, they have examined how people *think about* behaviour and norms; a second-order methodology.

Jonathan Haidt's [research](), in particular, suggests two very important insights. First, that normative statements are reflections of people's individual moral intuitions, not of an objective set of norms. Second, that people strongly *believe* that their normative statements refer to objective norms. Haidt found that people typically experience *moral dumbfounding* when all their arguments for a particular norm are refuted; they are unable to provide arguments for it, yet they refuse to abandon it. More importantly, they are *surprised* by this state of affairs.

It seems that people have a very strong tendency to believe that their normative statements are objective. This makes sense pragmatically; people who believe their normative statements are objective are likely to sound more convincing and less self-serving. The irony is that people fool themselves as well as other people.

This has some important consequences for meta-ethical questions in moral philosophy.

First, it's important to remember that philosophers are people too, and that the same self-delusion presumably applies to them as well. When philosophers advance normative theories, it's likely that not only are they advancing their personal moral intuitions, but that they *believe* that they're advancing objective normative truth. There's nothing special about philosophers in this regard; everyone believes that. This helps explain the reluctance to give up the belief in objective norms, even after most philosophers have given up God.

Second, that normative statements refer to a person's moral intuitions. This is not a new position, it dates back in some form to the ancient Greeks, typically referred to as *relativism*. However, with modern insights in evolutionary theory and cognitive science, it takes on a much more sophisticated form.

The standard model of relativism goes something like this: moral intuitions are like ice-cream preferences, some people like vanilla and some people like chocolate, and there's no objective norm dictating which flavour is better. This analogy is misleadingly simplistic, for several reasons.

First, flavour preferences are presented as arbitrary; there's no *reason* why someone should prefer one flavour over the other. This, however, is not true of moral intuitions. The [Moral Foundations Theory]() developed by Haidt and others, for example, explains moral intuitions as modules that developed because they provided evolutionary benefits, and whose function can be understood accordingly. This offers a powerful combination of objectivity and relativism. Moral intuitions are relative to an individual, but – because they evolved in all people as basic modules – are highly generalisable. The generalisability exists because of patterns of behaviour, not because norms are universal Platonic objects.

Second, flavour preferences are presented as singular decisions. This, also, is not true of moral intuitions. People have multiple intuitions that often overlap. Consequently, people are often choosing between *their own* intuitions, thus leading to a much more complex situation.

Third, flavour preferences are presented as causally closed; there are no consequences to choosing one flavour over another. This is not true of moral intuitions. Actions have consequences.

Fourth, flavour preferences are presented as fully known; people know what each flavour tastes like, and they know which one they prefer. Moral situations, however, often contain incomplete information. People not only don't always know the consequences of their actions, sometimes they don't even know their own preferences.

When adding in these elements, the model becomes a much better representation of moral discourse. People can share information about their different intuitions, as well as about facts in general. People can negotiate and/or deter people from certain choices, based on consequences. The four important aspects of moral discourse – signalling, debate, negotiation, and deterrence – don't require objective norms. All they require is a shared physical space where actions have consequences, and agents with incomplete information.

***

The implications of people accepting this – as I think they eventually must, as science continues to progress – is a change in moral discourse. Much as our discourse has gradually changed to become nontheistic, so must it gradually change to reflect the fact that people no longer accept objective norms. This means that instead of people carrying out moral discourse under the guise of theorising about objective norms, they simply engage explicitly in the four aspects of discourse mentioned above. This would surely make moral discourse more effective, in much the same way that it has become more effective without theism.

Similarly, the implication for moral philosophy is an abandonment of the search for objective norms. Instead of a top-down process governed by attempts to derive these norms through reason, it becomes a bottom-up process of observation and generalisation, just like any other science. People have moral intuitions; they are to some extent generalisable; they can be explained through non-moral theories such as evolutionary theory. There's no reason to think these intuitions are *entirely* generalisable into a single set of universal norms, but that's not necessary for effective science.