Empirical ignorance as defeating moral intuitions?
A puzzle for rule consequentialists (and others)

Rule consequentialists take an act to be wrong just in case it is forbidden by the best set of rules. And they understand what set of rules would be best in terms of the overall consequences of those rules being embedded in some sense in the community – widely or universally followed, on some views, or accepted, or internalised, by some or all or most of the community.[1] In defending this thesis, they tend to draw on the method of reflective equilibrium. They try to show that rule consequentialism is the best way of tying our more particular moral convictions into a coherent whole. This paper suggests that, if rule consequentialism is true, it's not possible to defend it in that way. The problem I'll describe won't be specific to rule consequentialists, but it will be especially sharp for them.

1 Introducing the argument

- Brad Hooker holds "an act is wrong if and only if it is forbidden by the code of rules whose internalisation by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being (with some priority for the worst off)" (Hooker 2000, 32)

- Derek Parfit holds that "everyone ought to follow principles whose universal acceptance would make things go best" (Parfit 2011, 375).

Both Hooker and Parfit use the method of reflective equilibrium. Hooker explicitly says: "the best argument for rule-consequentialism is that it does a better job than its rivals of matching and tying together our moral convictions, as well as offering us help with our moral disagreements and uncertainties" Hooker 2000, 104.[2] Parfit is similarly reliant on our particular

---

[1]Richard Brandt 1963, John Harsanyi 1977, Brad Hooker 2000, Derek Parfit 2011, Holly Smith 2010, and Michael Ridge 2006 describe a representative range of the options.

[2]The second advantage - offering us help with moral disagreements - is downstream from the first. It's only because the consequentialist theory matches and ties together our moral convictions that it's trustworthy enough to help us with our moral disagreements.

moral judgments. For example, he argues against Kantian strictures on treating agents as mere means from our intuitions about particular cases.

> THIRD EARTHQUAKE: You and your child are trapped in slowly collapsing wreckage, which threatens both your lives. Parfit 2011, 222. ... In Third Earthquake, you cannot save your child's life except by crushing Black's toe, without Black's consent. This act, I believe, would be justified Parfit 2011, 231

Parfit is here appealing to a very familiar kind of evidence: our moral intuitions, our considered moral judgments.

I'll argue that this familiar kind of evidence is actually off-limits for rule consequentialists, at least in many cases. It's off-limits because we're too ignorant about the consequences that would matter if rule consequentialism is true.[3]

I introduce my argument by giving an argument-schema, which you fill out by putting in a concrete intuition.

> IGNORANCE ARGUMENT, for a moral intuition that p:
>
> > 1. If I'm ignorant of the empirical facts that a normative theory takes to determine whether p, it's a mistake to take my intuition that p as evidence for that normative theory.
> > 2. I *am* ignorant of the empirical facts that the rule consequentialist takes to determine whether p.
> >
> > > (For example, I'm ignorant of empirical facts about the consequences of society-wide acceptance of certain rules.)
> >
> > 3. So it's a mistake to take my intuition that p as evidence for rule consequentialism if rule consequentialism is true.

The IGNORANCE ARGUMENT targets a particular intuition (schematically, the intuition that p) and concludes that the intuition that p is no evidence for rule consequentialism.

The bulk of this paper will defend Premiss One. After doing that, it will explore the scope of the IGNORANCE ARGUMENT: the range of moral intuitions where Premiss Two is plausible. I'll suggest that Premiss Two is

---

[3]So this argument is an extension of a challenge that James Lenman 2000 has made, about our cluelessness.

plausible for the key intuitions that the rule consequentialist uses to defend her view. So I'll conclude that, if Premiss One is true, it's impossible to defend rule consequentialism as the upshot of the method of reflective equilibrium.

2   Defending Premiss One

Imagine that I discover that some moral intuition is due to my socio-economic position. Every proponent of the method of reflective equilibrium would take me to have thereby discovered a compelling reason for not relying on that intuition in normative theorising.[4] I will call this sort of compelling reason for not relying on a moral intuition a *defeater*. So I will say that this case is a case where my socio-economic position is a *defeater* for my intuition.

I will argue that empirical ignorance patterns with my socio-economic position in this kind of case. Since we should see my socio-economic position as a defeater, we should also see empirical ignorance as a defeater.

As a warm-up, note that a defeated intuition can still be true. Even if I believe that your socio-economic position is a defeater for your intuition that you may ϕ, I can still believe that your intuition is true. So even though your socio-economic position creates some kind of defect in your epistemic position, we cannot say that it creates the defect of believing something false.

One natural idea is that intuitions due to my socio-economic position are defeated because they're due to morally irrelevant features of my epistemic position.[5] For example, the fact that I'm in socio-economic position $\mathcal{P}$ is irrelevant for what tax policy is best. But if my intuitions about tax policy are due to my being in $\mathcal{P}$, then my intuition would be due to morally irrelevant features of my epistemic position. And the intuition would be defeated for that reason.

I claim that empirical ignorance creates the same kind of defect that my socio-economic position creates. One way for moral intuitions to be due to my socio-economic position is for my *empirical* beliefs about tax policy to be due to my socio-economic position, and for my moral intuitions to be due to those empirical beliefs. In such cases, my moral intuition aren't due to the morally relevant features of tax policy, since the empirical facts are morally relevant and I'm not responsive to them in the right way. (One vivid illustration of this point is if my socio-economic position disposes me to have the same moral intuition even if I had different empirical beliefs.)

---

[4]For this kind of example, see TJ §9, p, 42.

[5]I'm grateful to [Referee 2 at *Analysis*] for suggesting this formulation.

This version of the socio-economic case exactly parallels cases where I have the moral intuition that p but I'm ignorant of the empirical facts that determine whether p. In both cases, my moral intuitions aren't due to the morally relevant features of the case in the right way. And in both cases, it would be a mistake to rely on the moral intuition. So Premiss One is true.

> 1. If I'm ignorant of the empirical facts that a normative theory takes to determine whether p, it's a mistake to take my intuition that p as evidence for that normative theory.

In general, then, my basic argument for Premiss One is that my socio-economic position is a companion in guilt for empirical ignorance. Since we take the former to be a defeater, we should take the latter to be a defeater as well.

3   Going deeper

A rule consequentialist who denies Premiss One will deny that empirical ignorance and socio-economic position are companions in guilt. And the most promising way for her to do that is to insist that empirical ignorance *doesn't* create any epistemic defect – that it's perfectly compatible with genuine moral knowledge. Then she can explain why my socio-economic position is different: since it undercuts moral knowledge, it does create an epistemic defect that empirical ignorance does not.

The rule consequentialist who makes this response is tacitly offering a diagnosis of why my socio-economic position is a defeater. She's assuming that my socio-economic position is a defeater for the intuition that p because it prevents me from *knowing* that p.

> (IMPOSSIBILITY OF KNOWLEDGE DIAGNOSIS) My epistemic position is a defeater for my intuition that p because no one in my epistemic position could know whether p.

And the DIAGNOSIS is available to a wide range of philosophers. It's available even to philosophers who insist that moral intuitions can be data for normative theorising even if we don't know them. For John Rawls 1971, 2005, for example, our moral beliefs can gain their epistemic standing in virtue of their relations to each other, whether or not they're initially known.[6] Such philosophers still can and should accept the IMPOSSIBILITY OF KNOWLEDGE DIAGNOSIS. The DIAGNOSIS is about cases where my

---

[6]For related discussion, see T. M. Scanlon 2002, 2014, and Norman Daniels 1979.

epistemic position makes knowledge *impossible*. And Rawls is making the claim that we can rely on intuitions that we *don't* know, not the claim that we can rely on intuitions that we *couldn't* know.

I agree that this Diagnosis is the best explanation of why my socio-economic position is a defeater. It allows us to focus on what should be the central question. Does empirical ignorance create the same kind of epistemic defect that my socio-economic position can create?

3.1 A hallmark of moral knowledge

The rest of this section will argue that empirical ignorance does create the same epistemic defect that my socio-economic position would create. Both undercut moral knowledge. They are companions in guilt. My argument will have two parts. The first part will describe a hallmark of what moral knowledge would involve if rule consequentialism is true. The second part will show that this hallmark is missing for anyone who's ignorant of the relevant empirical facts.

The hallmark I'll describe is a hallmark of knowledge in general, and not just of moral knowledge. Consider (S1).

> (S1) Anne knows that this glass contains water.

The hallmark of knowledge in general is that (S1) usually supports conditionals like (S2).

> (S2) If Anne gained testimonial knowledge that water is H_2O, she would be in a position to know that this glass contains H_2O.

(S1) supports (S2) because of general facts about what it takes to have knowledge. It's not specific to water or to H_2O.

Now the hallmark I'm describing is importantly defeasible. (S2) can sometimes be false while (S1) is true. Anne's learning about the nature of water could destroy her knowledge that this glass contains water. (Maybe she's convinced that the glass contains only atoms, not molecules, and would abandon her belief that the glass contains water if she learned about the nature of water.) Then (S1) would be true but (S2) false. So the hallmark is defeasible: it holds only when Anne's knowledge about the glass' contents would survive her learning about the nature of water.

We should see this hallmark is a general if defeasible feature of knowledge in general. So we should expect it to be a general if defeasible hallmark of *moral* knowledge, too. Consider (M1) and (M2).

(M1) I know that I may sacrifice someone's toe without their consent to save my child's life.

(M2) If I gained testimonial knowledge that rule consequentialism is true, I would be in a position to know that the rules that allow someone to sacrifice a toe without their consent to save anyone's life have the best consequences.

I conclude that (M2)'s being true is a general if defeasible hallmark of (M1)'s being true.

3.2 Empirical ignorance undermines this hallmark

I now show that this hallmark is missing for anyone who's ignorant of the relevant empirical facts.

Note first that conditionals like (M2) aren't trivial. Imagine both (i) that you're totally ignorant about what you ought to do in particular cases, and also (ii) totally ignorant of the consequences of different rules. If a reliable source of testimony told you that rule consequentialism is true, you still don't know what you ought to do. You also need to know about the consequences of different rules.

In other words, (M2) is true only if I'm in a position to be justified in believing something about the consequences of different rules. That is, I have to be *propositionally* justified in believing something about the consequences *conditional on* learning that rule consequentialism is true. Crucially, though, this conditionalized propositional justification needs to be grounded in categorical facts about me right now, facts that hold before I learn that rule consequentialism is true. If it didn't need to be grounded in that way, (M2) *would* be trivial and hold of everyone.

This section is addressing a rule consequentialist who insists that it's possible for me to know some moral claim even if I'm ignorant of the relevant empirical facts. At this point, we have enough tools to answer this consequentialist. If I know a moral claim, conditionals like (M2) will hold of me. And if I'm ignorant of the relevant empirical facts, conditionals like (M2) don't hold of me. So if I'm ignorant of the relevant empirical facts, I can't know the relevant moral claim. Empirical ignorance does generate the serious epistemic defect.

Now this argument assumes that if (M1) is true, (M2) is usually true. Why assume that, again? Parity of reasoning, from a general fact about knowledge. We saw that knowledge that this glass contains water usually supports knowledge that this glass contains H_2O conditional on learning that water is H_2O. Our initial, standard assumption should be that the

moral case patterns with the scientific case.[7]

The rule consequentialist might try one last ploy. I've admitted that my hallmark is defeasible, that it's absent if knowledge of rule consequentialism would destroy ordinary moral knowledge. So a rule consequentialist could in principle concede that this hallmark is absent but still insist that we do have moral knowledge. She just needs to hold that our ordinary moral knowledge would vanish if we learned that rule consequentialism is true.

This ploy tacitly concedes the crucial Premiss One.

> 1. If I'm ignorant of the empirical facts that a normative theory takes to determine whether p, it's a mistake to take my intuition that p as evidence for that normative theory.

If you're defending rule consequentialism, you're trying to show that we can come to know that rule consequentialism is true. If learning that rule consequentialism is true would destroy a piece of ordinary moral knowledge, then rule consequentialists would have to regard that piece of knowledge as defeated if she learned that her view is true. Then it would be a mistake to rely on that piece of knowledge in arguing for rule consequentialism.

It seems like the rule consequentialist has to accept Premiss One.

4   The scope of the Ignorance Argument

The IGNORANCE ARGUMENT targets a particular intuition (schematically, the intuition that p) and concludes that the intuition that p is no evidence for rule consequentialism. I've just defended Premiss One. I now turn to Premiss Two.

> IGNORANCE ARGUMENT:

---

[7]First: I'm not assuming anything controversial about closure principles. It's uncontroversial that (S2) is usually true if (S1) is, whatever your take on closure principles, since (S2) is a claim about what I'd *be in a position to* know, not what I actually know. Second: I'm not assuming anything controversial about the transmission of epistemic properties. Distinguish two claims:

> CORRELATION: If (M1) is true, (M2) is usually true.
> DEPENDENCE: The truth of (M1) usually depends on the truth of (M2).

Some philosophers grant conditionals like CORRELATION but deny claims like DEPENDENCE; James Pryor 2004 and Christopher Tucker 2010 are two examples. CORRELATION is enough to power the argument in this paper, so controversy about DEPENDENCE is irrelevant.

> ...
>
> 2. I am ignorant of the empirical facts that the rule
> consequentialist takes to determine whether p.
>
> ...

I think that the IGNORANCE ARGUMENT applies very broadly, and prevents rule consequentialists from defending their view as the outcome of reflective equilibrium. If rule consequentialists can't rely on intuitions where Premiss Two is plausible, they won't have enough intuitions left to defend their view.

Consider the THIRD EARTHQUAKE case. I think that I'm ignorant of the empirical facts that determine whether I'm permitted to sacrifice a toe without consent. So I think the IGNORANCE ARGUMENT applies to our intuitions about the THIRD EARTHQUAKE case. So rule consequentialists can't appeal to those intuitions as evidence for their view.

The method of reflective equilibrium is contrastive. You defend a theory by showing that it fits our undefeated intuitions better than the alternatives. So to defend rule consequentialism, you have to have enough *un*defeated intuitions left to defend the theory against the alternatives. And, crucially, the intuitions that uniquely support rule consequentialism tend to be even more subtle than our intuitions in the THIRD EARTHQUAKE case. So Premiss Two is even more plausible for those intuitions than for THIRD EARTHQUAKE. As a result, it's unlikely that we know enough to defend rule consequentialism as the outcome of reflective equilibrium.

Now one response to the IGNORANCE ARGUMENT is to insist that ordinary agents *are* in a position to know about the consequences that matter. There isn't space here to fully address this response. My main goal here has been to convince you that the rule consequentialist needs to care whether we're ignorant of the empirical facts. That's the point that the literature seems to have missed.

At the same time, though, it is odd to insist that we do know about the consequences. Imagine a character, *Ignorant Ian*, who we stipulate to be genuinely ignorant about the consequences. It's hard to articulate features of me that put me in a better position than Ignorant Ian. First: it's hard to find internalist, mental facts that distinguish us. For example, I haven't internalised any method for determining the consequences of different rules.[8] Second, it's hard to find external facts about me that distinguish

---

[8]Does Hooker's appeal to *expected* utility help? Maybe Ian, Parfit, and Hooker all are propositionally justified because propositions about the expected utility of the the rules are justified for them.

This suggestion requires implausible commitments in the metaphysics of

me from Ian. Now it is *more* plausible that external facts do distinguish us. We might, for example, appeal to the history of the population that I belong to. Maybe that population has converged on the rules with the best consequences. I'm propositionally justified in believing as I do in virtue of belonging to this population.

In general, though, this sort of external fact doesn't support propositional justification. Consider the proposition that I'm uniquely well-adapted to survive in my evolutionary niche. That proposition could well be true. But I'm not justified in believing it just because it's true. In order to be justified, the belief needs to be the upshot of a reliable process, or else grounded something more internal, like understanding or sense-data. We should treat the moral case similarly, absent special reasons to do otherwise. So we shouldn't take me to be justified in my beliefs just because I believe truly.

So in general, it looks like I am ignorant of lots of the empirical facts that the rule consequentialist takes to matter.

5 Wrapping up

I've introduced and defended what I called the Ignorance Argument. If the Argument is compelling, it's much harder for rule consequentialists to use the method of reflective equilibrium than has been appreciated.

---

morals. Suppose that the expected utility of the rules differs between Ignorant Ian and Brad Hooker. In order for this suggestion to help, what's wrong for Ian has to be different for what's wrong for Hooker; it's not enough if they just they have different *beliefs* about what's wrong. Ian has to be thinking about a property that is partially constituted by his credences and value function, and Hooker does too. If they're not thinking about different properties, it's hard to see how they could be justified in their beliefs in virtue of their credences and value functions. I don't count as justified in believing that there's a couch in front of me just because my credence that it's there is high enough; my credences only justify me in believing propositions about my (possibly idiosyncratic) epistemic state.

And it's implausible that Ian and Hooker are thinking about different properties. If they were, someone who knows all the facts will regard them as talking past each other if they disagree morally: Ian is talking about one property, and Hooker is talking about another. The moral plausible metaphysical claim is that they're disagreeing about a more objective property: maybe something connected to the expected value *for someone who is well-informed and reasons well.* But our original character, Ignorant Ian, could be genuinely ignorant about what has that property. My challenge is then to explain what distinguishes him from someone like me or Hooker.

Act consequentialists would also be vulnerable to the Ignorance Argument if they relied on the method of reflective equilibrium as much as rule consequentialists do. But they tend to be much *less* reliant on the method of reflective equilibrium. They ordinarily appeal to abstract intuitions about goodness and maximisation (Sidgwick 1907; Singer 1974; Hare 1973.) And the rule consequentialist can't plausibly appeal to those intuitions (Arneson 2005; Card 2007; Wall 2009). So because the act consequentialist can appeal to those abstract intuitions, cluelessness about the consequences matters less for her. She only needs to explain how an adequate decision procedure for maximising our chances of doing what we ought to do would vindicate our ordinary moral beliefs (Moore 1903; Smart 1973; Norcross 1990; Mason 2004; Railton 1984).

This paper has focused on rule consequentialists, because they need more from the method of reflective equilibrium. They need our ordinary beliefs to be *undefeated* evidence for their theory. As a result, our cluelessness is much more important for rule consequentialists than for act consequentialists. The rule consequentialist who recognises her own cluelessness can't defend her own theory.

But it's not just rule consequentialists who are vulnerable to the Ignorance Argument.

Ignorance Argument, for a normative view N

1. If I'm ignorant of the empirical facts that N takes to determine whether p, it's a mistake to take my intuition that p as evidence for N.

2. I *am* ignorant of the empirical facts that N takes to determine whether p.

3. So it's a mistake to take my intuition that p as evidence for N if N is true.

This argument is forceful against contractualist accounts of promise-keeping. Contractualists allow that you can reasonably reject particular principles about promise-keeping for their social effects, as Scanlon does (1998, 298ff). But it doesn't seem like ordinary agents like you or I are in a position to *know* about those social effects.

I've focused on rule consequentialists only because the Ignorance Argument is particularly troubling for them. But the Argument applies to any normative theory where normative facts depend on empirical facts that ordinary agents aren't in a position to know. If we accept it, we're pushed towards views where normative facts depend only on empirical facts that ordinary agents *can* know, as W. D. Ross (1930) thought.

Having introduced the IGNORANCE ARGUMENT, though, I want to end on a different note. I would like to find a way to resist the ARGUMENT. It doesn't seem like the right kind of reason to reject rule consequentialism. I think the mistake has to be back in the argument for Premiss One, in §3, where I identified a hallmark of moral knowledge. That hallmark is what powers the IGNORANCE ARGUMENT. I don't see how to resist the IGNORANCE ARGUMENT if we acknowledge that that hallmark is genuine.

So those of us who are sympathetic to rule consequentialism should think that there's a mistake somewhere in §3. But §3 reasoned by parity from structural features of scientific knowledge. Rule consequentialists should think that that reasoning is somehow mistaken, that moral knowledge is somehow different from scientific knowledge. But we owe everyone else an explanation why.[9][*]

References

Arneson, R. 2005. Sophisticated Rule Consequentialism: Some Simple Objections. *Philosophical Issues* 15: 235–251.

Brandt, R. 1963. Toward a Credible Form of Utilitarianism. In *Morality and the Language of Conduct*, eds. H.-N. Castañeda and G. Nakhnikian, 107–43. Detroit: Wayne State University.

Card, R. 2007. Inconsistency and the Theoretical Commitments of Hooker's Rule-Consequentialism. *Utilitas* 19(02): 243–258.

Daniels, N. 1979. Wide Reflective Equilibrium and Theory Acceptance in Ethics. *The Journal of Philosophy* 76 (5): 256–282.

Hare, R. M. 1973. Rawls' Theory of Justice. *The Philosophical Quarterly* 23: 144–155.

Harsanyi, J. 1977. Rule Utilitarianism and Decision Theory. *Erkenntnis* 11: 25–53.

---

[9] I do think it's possible to do that; for details, see Perl 2017.

Hooker, B. 2000. *Ideal Code, Real World: A Rule-consequentialist Theory of Morality.* Oxford University Press.

Lenman, J. 2000. Consequentialism and Cluelessness. *Philosophy and Public Affairs* 29 (4): 342–370.

Mason, E. 2004. Consequentialism and the Principle of Indifference. *Utilitas* 16 (3): 316–321.

Moore, G. E. 1903. *Principia Ethica.* Cambridge: Cambridge University Press.

Norcross, A. 1990. Consequentialism and the Unforeseeable Future. *Analysis* 50 (4): 253–256.

Parfit, D. 2011. *On What Matters*, Volume 1. Oxford: Oxford University Press.

Perl, C. 2017. *Positivist Realism.* Ph. D. thesis, University of Southern California.

Pryor, J. 2004. What's Wrong with Moore's Argument. *Philosophical Issues* 14: 349–77.

Railton, P. 1984. Alienation, Consequentialism, and the Demands of Morality. *Philosophy and Public Affairs* 13 (2): 134–171.

Rawls, J. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Rawls, J. 2005. *Political Liberalism.* New York: Columbia University Press.

Ridge, M. 2006. Introducing Variable-Rate Rule-Utilitariansim. *Philosophical Quarterly* 56: 242–53.

Ross, W. D. 1930. *The Right and the Good.* Oxford: Oxford University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other.* Cambridge, MA: Harvard University Press.

Scanlon, T. M. 2002. Rawls on Justification. In *The Cambridge Companion to Rawls*, ed. S. Freeman, 139–167. Cambridge University Press.

Scanlon, T. M. 2014. *Being Realistic about Reasons.* New York: Oxford University Press.

Sidgwick, H. 1907. *The Methods of Ethics.* Hackett.

Singer, P. 1974. Sidgwick and Reflective Equilibrium. *The Monist* 58: 490–517.

Smart, J. J. C. 1973. *Utilitarianism: For and Against.* New York: Cambridge University Press.

Smith, H. 2010. Measuring the Consequences of Rules. *Utilitas* 22: 413–33.

Tucker, C. 2010. When Transmission Fails. *The Philosophical Review* 119 (4): 497–529.

Wall, E. 2009. Hooker's Consequentialism and the Depth of Moral Experience. *Dialogue* 48 (2): 337–351.