Lecture 6: The Revisability Puzzle Revisited.

Puzzle from opening lecture: four somewhat plausible claims that are jointly inconsistent.

1. At any time, a person possesses a "highest level epistemic norm", that constitutes the person's standards of rational formation and retention of beliefs at that time.

2.(Assuming 1) It is not possible for the person to rationally revise that highest level epistemic norm under any conditions.

3. Any sufficiently high level rational norm must include a logic (powerful enough to be of use).

4. For any logic (that is powerful enough to be of any use), it would be possible for the person to rationally revise that logic under certain conditions.

#3 seems to have been basically supported by Lec. 2

#4 seems to have been basically supported by Lecs. 3&4.

I sketched prima facie arguments for Claims 1 & 2, in Lec 1. But they deserve a closer look—especially in light of the view about epistemology that I sketched in the previous lecture.

My view:

Claim 1 is questionable;

But the main problem is in Claim 2.

The idea of "a person's epistemic norms" is ambiguous (even when the notion of norm itself is understood univocally).

It could mean

the epistemic norms that he *is committed to*,

or the epistemic norms that he *employs in making epistemic evaluations*,

or the epistemic norms that he *employs in forming and retaining beliefs*.

My claim: in *any* of these senses, a person's norms are subject to rational change, no matter how "high level" they may be.

It is the third sense of "a person's norms" for which the possibility of rational change is most controversial.

That will be my ultimate focus. But understanding rational change of epistemic norm in the third sense requires attention to rational change of norm in the other senses.

A further clarification of "a person's norms" in the third sense:

I'm taking norms (at least, the ones that are the best candidates for being unchangeable) as *policies* or *rules* of some sort.

But they aren't the kind of policies or rules that the agent needs to have explicitly represented in his head.

> That would lead to familiar regress, in that we'd need policies or rules for processing the internal representations.

Rather: these are policies or rules that are *implicit in the person's practice*.

What does this amount to?

To attribute an epistemic policy or rule of this sort is to give an *idealized description* of how the person forms and alters beliefs.

For a variety of reasons, *there need be no best* idealized description.

> 1. There are different *degrees of idealization*: for instance, some idealizations take more account of memory limitations or computational limitations than do others.
>
> 2. (Probably less important:) There can be multiple good idealized description with the same degree of idealization, especially when that degree of idealization is high:
>
> > Since a description at a given highly idealized level only connects loosely with the actual facts, there is no reason to think it uniquely determined by the facts.

There are *multiple candidates* for the best idealization of our epistemic behavior.

> Any such idealization counts any factors it doesn't take into account as non-rational.
>
> Insofar as the idealization is a good one, it is *appropriate* to take the factors it doesn't take into account as non-rational.

The lack of a uniquely best candidate for one's basic rule is largely due to a lack of a uniquely best division between rational and non-rational factors.

This doesn't itself go against the idea of a highest level norm guiding our behavior.

> Since in attributing norms one is idealizing, the issue of a highest level norm is the issue of whether *a good idealization will postulate* a highest level norm.
>
> Doing so is *compatible with different good idealizations postulating different highest level norms* (whether of "the same level" or "of different levels").

Indeed: the *prima facie* implausibility in the idea of a "highest epistemic norm" governing our behavior is *lessened* by the fact that it doesn't entail that there's a uniquely best candidate for what this norm is.

Still: why should we believe that the best idealizations will posit a highest level norm?

Here's the answer that I gave in "Apriority as an evaluative notion" (2000). (I sketched a similar answer in the opening lecture, without endorsing it.)

> "...the alternative is an idealization that postulates multiple norms, each assessable using the others.
>
> But there is an obvious weakness in an idealization of this alternative sort: it is completely uninformative about what the agent does when the norms conflict.
>
> There is in fact some process that the agent will use to deal with such conflicts. Because this conflict-breaking process is such an important part of how the agent operates, it's natural to consider it part of a norm that the agent is following. If so, it would seem to be included in a basic or highest-level norm, with the "multiple norms" really just default norms that operate only when they don't come into conflict with other default norms.
>
> Of course, the process of resolving conflicts provided by this basic norm needn't be deterministic; and as stressed before, there need be no uniquely best candidate for what the higher norm that governs conflict-resolution is. But what seems to be the case is that idealizations that posit a basic norm are more informative than those that don't."

In retrospect, it's hard for me to see the force in this. Two problems.

Problem 1:

Since there are different degrees of idealization, why shouldn't the process that decides the conflict among norms at one degree of idealization be excluded from those norms, but included only in a norm of lower degree of idealization (i.e. one that takes more account of computational structure)?

Problem 2:

I think there's a problem in going from the claim that

> the resolution-breaking process is intuitively rational

to the claim that

> the resolution-breaking process is included in the agent's norms.

This is closely related to the main problem I'll find with the argument for Claim 2, so I won't go into it here. (It isn't needed: Problem 1 is enough to make Claim 1 unobvious.)

Claim 2 was: *If* a person is guided by a highest level epistemic norm, it isn't possible for her to rationally revise it under any conditions.

The *prima facie* rationale given in Lecture 1:

(a) Rational revision requires the use of a norm (one that declares the revision rational);

(b) If the rational revision of a norm N that one employs went by the use of some norm other than N, N couldn't be a "highest level" norm that one employs;

and (c) No norm can dictate its own revision.

I'll try to sketch a believable account of norm-change that gets around this. Which premise(s) does it undermine? The account will reveal some ambiguities in them, and so I'm not sure any are entirely free of blame. But I'd say that most of the blame lies with (a).

But I'll begin by discussing (c).

The *typical* mode of rational change in the norms one employs in forming and retaining beliefs is a two-stage affair:

> Stage 1: keeping the norms one employs in forming and retaining beliefs constant, one modifies *the norms that one is committed to and/or that one employs in evaluations*;
>
> Stage 2: one then brings the norms that one employs in forming and retaining beliefs into line with these new commitments and/or evaluations.

(c) was:

> No norm (that an agent could follow) can dictate its own revision.

The thought behind it was:

> How could any norms—or any ones that an agent could follow—tell us to revise themselves? Wouldn't following those norms require not following them? This seems incoherent, or at least to make following the rule impossible.

But the problem here seems to be only in Stage 2; *this thought* raises no obvious problem for the use of the norm to "undermine itself" in the sense of Stage 1. (I'll later discuss an additional worry about that.)

More fully:

The thought behind (c) doesn't rule out that by following a highest level norm N, we could be rationally led *to conclude that we shouldn't be following N, and should instead be following an alternative N**. (Stage 1)

It only tells us that N wouldn't then dictate *switching from N to N**. (Stage 2: the switch proper)

That doesn't mean that we wouldn't make the switch—only that in making it, we wouldn't be following N.

It also doesn't mean that we wouldn't be *rational* in making the switch—only that were we to rationally make it, *the rationality would not be explained by its being in accordance with norm N.*

The argument for (c) *does* show that there's no hope of explaining the rationality of Stage 2 of the transition from N to N* by means of N.

What *could* explain the rationality of Stage 2?

One suggestion: its being in accord with some other norm. That would be denying (b).

My objection to this in Lecture 1: it doesn't fit with the hypothesis that N is a "highest level" norm I'm following.

But here too we need to be more careful.

The clear truth behind (b): Suppose that on a certain degree of idealization of my behavior, N is the highest level norm I'm following. Then:

The rationality of the switch from N to N* (the switch proper, i.e. Stage 2) can't be explained by its being in accord with any norm *that I'm following in making the switch*. Or at least, not any that I'm following *according to the degree of idealization in question*.

If other norms are relevant to the rationality of the switch proper, they must *either* be

(i) norms at a different level of idealization,

or

(ii) norms that I wasn't following at all.

It's not out of the question to try to explain the rationality of the switch proper in terms of (i) or (ii) (thereby perhaps falsifying (b) on one interpretation). [For (ii), the idea would be to use something like "the correct norm" to explain it.]

But I think it's better to say that *we don't need to invoke norms at all to explain the rationality of the switch proper*. So *on the question is the rationality of the switch proper,* i.e. of Stage 2, *I'll take (a) to be the main culprit.*

But first, a long digression, on Stage 1 (where I'll also argue that (a) is to be rejected).

To repeat: *The "thought behind (c)"* does nothing to rule out our using our most fundamental norm at the preliminary stage, Stage 1.

> That is: it does nothing to rule out that by following a highest level norm N, we could be rationally led to conclude that we shouldn't be following N, but should instead be following an alternative N*.
>
> It only tells us that N wouldn't then dictate the switch from N to N*, and that if that switch is rational, we need to explain it's being so in some other way.

But maybe *something else* rules this out?

> That is: rules out the possibility of following *one* highest level norm and concluding on its basis that one *should* be following *a different one*?

Results on **"immodest inductive methods"** (Lewis) purport to show that *each inductive method predicts that it will do better than all the alternatives to it*.

These results are overstated.

> Attempts to prove that every method declares itself better than all others are based on simplistic 1-dimensional *criteria of what's good in a method*.
>
> And even given the criteria for what's good, the arguments depend on controversial *measures of closeness to the good*. Gibbard has shown that alternative measures of closeness do *not* yield the result that all methods declare themselves best.

Still, it's hard to take much comfort in the Gibbard results:

> He provides alternative measures of closeness, on which there are methods that don't declare themselves best. But these measures are ones for which *the only methods that do declare themselves best are exceptionally bad ones!*
>
> There's no reason to think that there is a way to spell out a criterion of betterness (i.e. of goodness and of closeness to the good) on which most methods that ought to come out bad will *declare themselves* to be bad, but on which methods that ought to come out good will declare themselves good.

So: while the problem of immodesty is overstated, **there does seem to be a deep underlying problem with the idea that rational debate between consistent advocates of alternative inductive methods would lead to the better method winning out over the worse.**

And it isn't implausible to extend this beyond inductive methods to "highest level norms" more generally.

I grant this.

But it's compatible with

> **rational debate about norms** (including "highest level ones", if such there be),

and

> **rational change in norms, emanating from such debate.**

Rational debate about norms. (Stage 0: a precursor to change even in which norms are advocated.)

To get a model for how debate about any deeply entrenched belief or norm proceeds, it's important not to rely on an idealization according to which agents are logically omniscient.

Actual agents don't see all the consequences of their beliefs, policies and preferences.

This *failure of logical closure* leads also to many unrecognized *inconsistencies*.

In the previous lecture I talked of the cognitive state of an agent as represented by

> a "pure component" that consists of a measure on a set of worlds (or a set of such measures),

and

> a normative component that consists of a precise norm (or a set of precise norms, or a measure on the precise norms, or a set of such measures, or ...).

This is an acceptable idealization in circumstances where inconsistencies play no role. But **in contexts where inconsistencies *do* play a role, it blinds us to how rational debate proceeds**.

Outline of a better picture:

In the case of pure belief:

Suppose that at any time, an agent X has a certain body of *core doxastic attitudes* toward non-evaluative claims.

The exact form of these core doxastic attitudes is a matter for psychology.

On a degree of belief psychology, it might include:

> specific degrees of belief for certain claims;
>
> specific degrees of conditional belief (like one's degree of belief that heads will result *given that the coin is flipped*);
>
> upper and lower bounds on degree of conditional belief;
>
> comparative degree of conditional belief ("*A* is more likely given *C* than *B* is given *D*")
>
> etc.

This set of attitudes will *not* be deductively closed (or closed under probabilistic consequence): one has no degree of belief in very complicated logical truths that one hasn't explicitly contemplated.

And it is unlikely to be consistent or probabilistically coherent: e.g. an agent might well have a degree of belief less than 1 in some complicated logical truth.

One might invent methods for assigning to such an agent a set of probability functions to represent the pure doxastic state.

> (Rough idea: take the agent's pure doxastic state to be the set of probability functions that satisfy *sufficiently many* of the agent's core doxastic attitudes.)

But it isn't clear how to satisfactorily work out the details, and the result would obscure important features of rational debate more than it would illuminate them.

Extend this to impure doxastic attitudes:

> either directly (attributing to the agent a certain body of *core impure doxastic attitudes*, toward *non-evaluative and evaluative claims together*);
>
> or indirectly, attributing to the agent a body of core policy-commitments, core preferential commitments, etc., which together with the beliefs will generate the impure attitudes.

Either way, the set of attitudes won't be deductively closed or consistent.

Again, one can invent methods for assigning to an agent subjective measures μ and υ on the worlds and norms respectively, or rather, sets of pairs of such measures. (The idea, as before, would be to look at sets of pairs <μ,υ> that satisfy sufficiently large subsets of the agent's impure core.)

But again, an understanding of the dynamics of doxastic states is better achieved without this.

The reason (as in the case of pure doxastic attitudes): **The measures are an epiphenomenon: the real work goes on at the level of the core attitudes.**

If the core attitudes were consistent and evolved in accord with a very demanding picture of "idealized rationality", the measures (or sets of pairs of measures) would evolve in a smooth way that could be described without mention of the underlying core.

But since the core attitudes aren't even consistent, the evolution of the measures won't be characterizable without reference to the underlying core.

Even without a detailed account of how these inconsistent cores evolve, we can see how this picture opens up the possibilities for rational normative debate.

The point is obvious: in debate, **one (consciously or unconsciously) exploits inconsistencies and other tensions in the other person's views**.

> That's true even in non-normative debate. In normative debate, the point is expanded: not just tensions within the normative commitments, but also tensions between the normative commitments and the norms they employ in acting/believing or in evaluating.

(By a tension in one's views, I mean an uncomfortable commitment: e.g. an unseen commitment to something the agent wouldn't accept were it made explicit.)

In convincing someone of a claim *A*, we typically argue to it from things the person explicitly accepts and other things he can be easily brought to accept.

The person may resist the argument, by questioning some of the claims used in it (even ones he previously accepted); but a good arguer is likely to find other ways to argue for *A* from things he accepts. With enough such argument the person is likely to be persuaded to alter his views and accept *A*.

If the person had debated someone else, he might well have been led to resolve the inconsistency in his views by keeping *A* and altering some of the related views.

This process of revision can lead to fundamental change in both pure and impure core attitudes.

In the latter case, if the overall change is important enough, it will constitute **a change in the norm advocated**. (Which norm one advocates is determined by which norm provides the best fit for one's doubtless-inconsistent set of impure doxastic attitudes).

But will such a change be rational?

Stage 1: Under what conditions will a change in view produced by rational debate be rational? (In particular, a change in the norms one advocates.)

Rational debate (even among parties acting in good faith) can lead a person to replace a better view by a worse one. In which such cases is the *change* rational?

> There are cases of this sort where *I'd be inclined* to call the change rational. (Primarily, cases where the argument is extraordinarily compelling, and the resultant position not a whole lot worse than what it replaced.)
>
> There are other cases where *I'd be much less inclined* to call it rational. (Primarily cases where the faulty norm is much worse than what it replaced, and where the convinced party may have had some grounds for suspicion.)
>
> There are cases where I feel conflicting inclinations: for instance, someone convinced by a highly persuasive but faulty argument for adopting a statistical procedure that in fact is deeply flawed.

Do we need a theory of when such changes are rational and when they aren't, that will decide such cases?

On what in the last lecture I called "the dipstick model", I guess we do need such a theory: we need to know how much "epistemic fluid" is produced by a persuasive but faulty argument, and how much of this is then pumped out by any "rational intuitions" against the conclusion.

But on the view I proposed, there is nothing objective to measure. To call a change rational is to express (a certain kind of) approval. And approval and disapproval are *multi-faceted*.

Suppose

> I see that Jones's initial statistical procedure, before being convinced by Smith's incorrect argument to change it, was a good one. And I know how Smith's argument went wrong.
>
> But I also see the apparent power of Smith's argument to anyone not immersed in subtle issues in the philosophy of statistical inference.

Then

> I positively evaluate Jones's intellectual honesty in following out reasoning that seems persuasive, and being willing to revise his work in light of this. I minimize the fact that he erred, since the fallacies involved were subtle.
>
> At the same time, the costs of employing Smith's faulty statistical procedure may be fairly high, so I think there is also something negative about Jones's conversion.

Why think that more needs to be said? Why think there needs to be a single standard of reasonableness, and that these two factors need to be weighed against each other?

I should regard Jones as in an unsatisfactory *overall* credal state.

But *weighing each belief in the state on its own, and on a single scale, is pointless*.

I take the upshot of the above to be:

> Rational change *even as to the norms one advocates* needn't be entirely norm-driven.

Moreover:

> To the extent that it *is* norm-driven, it needn't go entirely by *consistent use of* one's highest-level norm (supposing there to be such a thing): failures of logical closure and of consistency play a crucial role.

So in the argument for Claim 2, assumption (a) ("rational change requires the use of a norm that declares the revision rational") would be incorrect *even for Stage 1 of the revision process*, i.e. the stage concerning which norms *to advocate*.

Stage 2: These remarks apply not only to change in the norms one advocates, but also to change in the norms that guide one.

The usual way to change the norms that guide one is to first come to advocate new ones and to then train ourselves to act or reason in terms of them.

When is the last step rational?

This I take to not be a straightforwardly factual question, but a question of evaluation.

There are cases where I'd be more inclined to evaluate the *advocacy* of a change in norms as rational than to evaluate the *employment* of the change of norms as rational.

> E.g. cases where the person is seduced by persuasive arguments for skepticism?

There are also cases where a person may change the norms that guides him not as a result of rational argument for changing the norms he advocates, but in some other way.

The most important other way: the change in guiding-norm is still a result of a change in advocated norm, *but the latter change isn't due to rational debate but, say, to change in preference*.

Example: we might come to prefer inductive policies that self-correct more slowly, or that are more cautious about accepting generalizations; and such changes of preference might lead us to "retrain ourselves" to follow inductive policies more in accord with these new preferences.

> (Recall: the inductive policies are not literally goal-driven, so there is no obvious argument that such a change in inductive policy would be guided by a more fundamental norm.)

In these cases, I'd probably count the change as rational if the new guiding norm is far superior to the old one.

But again, **there is really no issue worth debate in these cases: it's a matter of evaluation, not of metaphysical fact.**

Here too, assumption (a) ("rational change requires the use of a norm that declares the revision rational") is just incorrect.

***

This completes the general discussion of the puzzle that these lectures have been concerned with.

But a specific sort of change of norm has occupied a special place in the lectures: viz., change of logic. To conclude, I relate the general discussion to this special case.

Example:

Phase 1: Joe starts out with a norm that allows reasoning in accord with **classical logic, together with the Intersubstitutivity Principle**, i.e. the equivalence of True(<*A*>) to *A*. His norms are inconsistent. (Indeed, they're trivial: they imply everything.)

Phase 2: On being shown the inconsistency, his first move is to adopt a **gap theory, or some other theory that gives up the Intersubstitutivity Principle while retaining classical logic**. He learns to operate with this system: these are his new norms.

Phase 3: Further discussion leads him to think that this wasn't the best way to go: he should have kept the **equivalence of True(<*A*>) to *A*, while weakening his logic in a certain way.** He then learns to do this: another shift in his basic norms of reasoning.

How do these shifts happen? Typically, the process starts by rational debate. (Not necessarily debate with someone else: maybe just internal conflict within oneself, that one thinks through.)

The first shift in norm, from inconsistent (indeed, trivial) norms to something else, is easy to motivate: a norm that leads to any conclusion at all isn't very useful.

But how does the person settles on a consistent replacement? (Esp. if he knows of more than one, and has to choose between them; or if he's deciding which type of replacement to try to work out.)

Typically one will look at the different replacements one thinks of or is told about, and think as best one can about their consequences and what it would be like to live with them, and on this basis make a choice. The details will depend on happenstance and on the agent's psychology.

The choice won't be made via one's earlier norm—that was trivial, after all.

It may be guided by *portions* of that trivial norm, but different portions could have ruled differently. Very likely, the final transition won't be based on any process that is deterministic at the psychological level.

What makes it rational or irrational?

There's no hidden fact here: in this case, we'll mostly be inclined to call it rational to the extent that we approve of (i) *the process that led to it*, but to some extent our judgement of (ii) *the view itself* will also enter in.

It's similar for **the second shift in norm**, say from a consistent gap theory in classical logic to a consistent view that retains Intersubstitutivity in a non-classical logic.

Such a shift is likely to be produced by tensions within one's credal state: consequences of one's views that one isn't happy with. These consequences can lead one to try out an alternative theory.

> E.g., we may notice that many of the standard paradoxes result from regarding $A \leftrightarrow \neg A$ as equivalent to $\neg(A \leftrightarrow A)$ and hence inconsistent. We know that's part of classical logic, but at the same time see the possible explanatory benefit of rejecting it.
>
> We think about what this would involve; also, about the fact that not all paradoxes turn on it, so we'd need to generalize to accommodate others. We look at different attempts to get a general theory, notice their limitations, try to improve on them, notice costs of so doing, etc..
>
> Perhaps we draw on analogies between the semantic paradoxes and quasi-paradoxes of vagueness.
>
> (The chance of a transition is greatly increased if one is engaged in conversation with an advocate of the non-classical view who is adept at arguing for it.)

The change won't be simply a product of the prior norm, because it is inconsistent with that norm.

Again, whether and when one makes the change is a matter of individual psychology, and probably not governed by deterministic laws at the psychological level.

On the question of when the change is rational, we consider several things:

> We look at the details of the process that led to the revision. (E.g., did the person think carefully through the difficulties with the old norms and think through what's involved in using the new ones?)
>
> We make our own judgement of the merits of the norms he ended up with.

From these, we make a multi-faceted evaluation of the ways in which we approve and the ways in which we don't.

And that's all there is to it: there is no factual question about rationality that we've left out.