

Ethics and probability

Frank Jackson

Seminar in PHI/CHV 524

24 October 2012

What I'll do

- Talk about a famous kind of example and see what it might teach us.
- About 2/3 way through we'll note some implications for the Procrastinate examples (here you may find the extract from Doug Portmore useful).

An example to fix our thinking

- A bomb is set to go off shortly. Joe's choices are:
 - enter 1 on the bomb's control panel
 - ...
 - enter 10^6 on the bomb's control panel
 - evacuate the building without entering a number
- Joe knows
 - one number is the code to defuse the bomb; entering any other number will set the bomb off and many will die.
 - there is no way to find out which number defuses the bomb (though in fact it is 17).
 - if he evacuates the building, it will be destroyed and people will be injured but no-one will die.

What ought Joe to do?

- This much is obvious: there are two viable candidates.
 1. What Joe ought to do is enter 17 on the bomb's control panel.
 2. What Joe ought to do is evacuate the building.
- Any other answer is for sure wrong.

How things look on one style of non-consequentialist theory

- *Noah's rocket examples* as objections to (traditional) consequentialism
- Earth is doomed; you, as one of those to be left behind, have to decide who goes onto the rocket to New Earth; it is down to the last 11 places; the candidates are Mary, Dick, Ann, Harry,
- Both Mary and Dick would be valuable additions but Mary would be a bit better; Ann, Harry, ... are trouble makers.
- Everyone on the rocket gets a drug that erases any memories of how they came to be selected.

- Dick is your only child/the person you promised a place to/the person who saved your life/the person who alerted Earth to the problem in the first place/the person who has lived the most valuable life so far/the person who has had much the worst time up to now/— .
- Anti-consequentialist claim: for some /—, you ought to give the last place to Dick ahead of Mary.
- The erasing drug means that consequentialists cannot reply by noting the good effects of rewarding good behavior, cementing family ties, etc.
- Non-consequentialists typically grant that if Mary is enough ahead, that outweighs /— .

Bringing probability into the story

- You know Mary isn't /—, and that exactly one of Dick, Ann, Harry, ... is /— but don't know which.
- You know that Mary would be a bit better than whoever it is who is /—.
- You know the rest are troublemakers.
- For each of Dick, Ann, Harry, ..., (everyone minus Mary), there is a 10% chance that they are /—, and a 90% chance that they are a trouble maker.
- It is Dick who is in fact /—, unknown to you.
- What should non-consequentialists say you ought to do?
- There are two viable candidates: select Mary, select Dick.

Back to the bomb example

- As we said before, this much is obvious: there are only two viable two candidates for what Joe ought to do:
 - enter 17 on the bomb's control panel
 - evacuate the building.
- Any other answer is wrong.
- An obvious first up reaction is to say that there are really two questions here; this allows us to say that both answers are correct answers to their respective questions.

- What Joe ought *objectively* to do is enter 17 on the bomb's control panel.
- What Joe ought *subjectively* to do is evacuate the building.
- Our options as philosophers
 - Two concepts, separate but equal
 - Two concepts, with the objective one basic
 - Two concepts, with the subjective one basic
 - One concept properly speaking, the subjective one.
 - One concept properly speaking, the objective one.

- The thought behind the demand for a single answer.
- Saying that there two answers, albeit to two different questions, is a classic case of ducking an important question.
- "Come clean and tell us what Joe ought to do: he's waiting for answer!"
- (Remember the debate over Newcomb's paradox.)
- What I said (and many say and many deny): well, really, what Joe ought to do is to evacuate the building.

- A moral theory worth the name should deliver the answer to what a subject ought to do, on some given occasion, in the situation they find themselves in. Isn't moral theory, at the end of the day, concerned with decision, and any decision is necessarily made in an epistemic context?
- What's more, we don't want to divorce what Joe ought to do unduly from what it is rational for Joe to do in decision theory terms.

- Why give the "objective" ought the time of day?
- One reason might be the desire to retain the link between what ought to be done and what an agent has most reason to do, combined with a 'externalist' view of when an agent has a reason.
- In some good sense, there exists a reason for Joe to enter 17 on the bomb's panel.
- (If there's a tiger in the next room, there is a good reason for me not to enter the next room even if I am unaware of the tiger.)
- I think this line of thought rests on an equivocation.

- There are two ways to read 'having a reason'.
- On one, there is a reason to phi when phi-ing would achieve something good even if the agent is quite unaware of the fact. This is the sense of 'having a reason' on which it is true that Joe has a reason to enter 17.
- On the other, there is only a reason if the agent is aware of the fact.
- Only on the second is there an intimate link between what an agent has reason to do and what an agent ought to do.

- So again we might ask, Why give the "objective" ought the time of day?
- We aren't playing "no speaks" (*Tractatus*: What we cannot speak about we must pass over in silence.)

- You are right, there has to be one preferred answer and it is the one you went for, but there's a problem that needs discussion.
- Consider: 'If x is the right number, Joe ought to enter x on the control panel'
- Evidentially, it is true for every value of x
- It has a true antecedent for $x = 17$ (unknown to Joe).
- *Modus Ponens* then delivers the answer that Joe ought to enter 17 on the control panel.
- We have a CE to Modus Ponens!

- Conditionals are being confused with conditional probabilities.
- Anyone who holds that evacuating the building is the correct answer for what Joe ought to do should hold that

If x is the right number, Joe ought to enter x on the control panel

is false when $x = 17$.

- However, anyone who holds that evacuating the building is what Joe ought to do is allowing that what Joe ought to do is a function of his credence function, and if we conditionalise his credence function on ' x is the right number', what he ought to do is enter x on the control panel. That's what is true for all x .

- Agreed, there is one answer for what an agent (really) ought to do on any given occasion (setting aside indeterminacy and ties) but you made the wrong choice. It's the "objective" one.
- What Joe ought to do is enter 17 on the control panel – and who said that Joe, or anyone, always knows what they ought to do?
- But now we need words to describe the option of evacuating the building.
- Unless we want to publish a paper with a title like "In praise of no speaks".

- Easy: the option of evacuating the building is the option that it is rational for Joe to hold is what he ought to do given what he knows at the time.
- But it isn't if the supporters of the objective ought being the one true ought are right!
- Joe knows for sure that evacuating the building isn't the option that is objectively right.
- What's subjectively right \neq what's probably objectively right. Indeed, we should replace the term 'subjectively right' with 'expectably right' – ugh.
- Is evacuating the building the option we should praise?
- Not necessarily; praising may have bad effects.

The double standard worry

- You in effect idealise Joe's desires but fail to idealise his beliefs. Be consistent!
- But Joe's beliefs are subject to some constraints. He is required to seek more information unless it costs too much.
- Morality is concerned with value, and that connects most closely with what we ought to desire, not with what we ought to believe. Some asymmetry is to be expected.
- (A sort of reversal of Hume's famous remark about scratches.)

A too cheap way out

- Could we have described Joe's options earlier as:
 - evacuate the building, versus
 - insert the right number into the control panel?
- This is theft over honest toil!
- We need to think in terms of what we are certain we can bring about (which will typically be very limited), combined what is likely given each of the things we are certain we can bring about.
- How it looks for evacuating the building: Joe is certain he can shout "Evacuate the building". What then comes into play is how likely that is to lead to the building being evacuated, and so on.

The Procrastinate example

- Prof. P is invited in January to review a book; the review is due July 1.
- The ranking of possible situations from best to worst is:
- Says yes in January, delivers by July 1. (Very good)
- Says no. (Not too bad; someone else will do the job but they won't be as good.)
- Says yes but fails to deliver. (Very bad, book goes unreviewed, author misses out on tenure.)
- Although the course of action of saying yes in January and delivering by July 1 is one that Prof. P is able to follow, in fact he won't follow it.
- The original version makes no mention of probability.

A possible confusion

- You might be attracted by both:
- i) If it is true at t_1 that S ought to do A at t_2 , S is able at t_1 to form an intention to do A at t_2 .
- ii) It is not possible to form an intention at t_1 to do A at t_2 if you believe you will not do A at t_2 .
- And that the combination of these two principles means it isn't true that Procrastinate should say yes in January and review by 1 July. Despite the fact he could say yes and review by 1 July, the fact that he won't means that he couldn't form the needed intention.
- But the *fact* that he won't doesn't mean he couldn't form the intention, and what he believes is irrelevant to the question about what he objectively ought to do.

What happens if we frame the Procrastinate case in terms of what's probable?

- Replace 'If he says yes, he won't review by 1 July' by 'It is *probable to degree x* that he won't review by 1 July given he says yes'.
- Make saying yes and not reviewing bad enough to mean having a probability of $x < 1$ of happening is enough to make the option of saying yes in January worse than saying no. (This is possible even if x is small.)
- But now the only principle linking what you can intend with what you believe that might be relevant would be:
- It is not possible to form an intention at t_1 to do A at t_2 if you believe there is some chance you will not do A at t_2 .
- No-one accepts that principle.

Some literature

- Frank Jackson and Michael Smith, "Absolutist Moral Theories and Uncertainty", *Journal of Philosophy*, CIII (2006): 267–283
- Michael Zimmerman, *Living with Uncertainty*, CUP, 2008
- Michael Huemer, "Lexical priority and the problem of risk", *Pacific Philosophical Quarterly*, 91 (2010): 332-351.