Reading list for weeks 1, 2, 5, 6, 8

Week 1: Consequentializing

Core reading (mandatory)

- Portmore, Douglas W. "Consequentializing." *Philosophy Compass* 4.2 (2009): 329-47. Link.
- Brown, Campbell. "Consequentialize This." *Ethics*, vol. 121, no. 4, 2011, pp. 749–771. Link.
- Schroeder, Mark. "Not So Promising After All: Evaluator-Relative Teleology and Common-Sense Morality." *Pacific Philosophical Quarterly* 87.3 (2006): 348-56. Link.

Further reading (optional)

- Portmore, Douglas W. "Combining Teleological Ethics With Evaluator Relativism: A Promising Result." *Pacific Philosophical Quarterly* 86.1 (2005): 95-113. Link.
- Portmore, Douglas W. "Consequentializing Moral Theories." *Pacific Philosophical Quarterly* 88.1 (2007): 39-73. Link.
- Dreier, J. "Structures Of Normative Theories + Addressing The Agent-Neutral And Agent-Centered Distinctions As Substantive Mistakes In Moral Theorizing." *Monist* 76.1 (1993): 22-40. Link.
- Louise, Jennie. "Relativity of Value and the Consequentialist Umbrella." *The Philosophical Quarterly (1950-)* 54.217 (2004): 518-36. Link.
- Vallentyne, Peter. "Gimmicky Representations Of Moral Theories." *Metaphilosophy* 19 (1988): 253–263. Link.

Week 2: Incomparability and parity

Core reading (mandatory)

- Chang, Ruth. "The Possibility of Parity." *Ethics* 112.4 (2002): 659-88. Link.
- Hsieh, Nien-Hê. "Is Incomparability A Problem For Anyone?" *Economics and Philosophy* 23.1 (2007): 65-80. Link.
- Rabinowicz, Wlodek. "Value Relations." *Theoria* 74 (2008): 18–49. Link.

Further reading (optional)

- Broome, John. "Is incommensurability vagueness?" in Chang, Ruth. *Incommensurability, Incomparability, and Practical Reason*. Cambridge, Mass.; London: Harvard UP, 1997. [not online] Link.
- Rabinowicz, Wlodek. "Incommensurability and Vagueness." *Proceedings of the Aristotelian Society, Supplementary Volumes* 83 (2009): 71-94. Link.
- Hare, Caspar. "Take the Sugar." *Analysis* 70.2 (2010): 237-47. Link.
- Schoenfield, Miriam. "Decision Making In The Face Of Parity." *Philosophical Perspectives* 28.1 (2014): 263-77. Link.
- Gert, Joshua. "Value and Parity." *Ethics* 114.3 (2004): 492-510. Link.

- Chang, Ruth. *Incommensurability, Incomparability, and Practical Reason*. Cambridge, Mass.; London: Harvard UP, 1997. [not online] Link.

Week 5: Actualism and possibilism

Core reading (mandatory)

- Jackson, Frank, and Robert Pargetter. "Oughts, Options, and Actualism." *The Philosophical Review* 95.2 (1986): 233-55. Link.
- Ross, Jacob. "Actualism, possibilism, and beyond" in ed. Mark Timmons. *Oxford Studies in Normative Ethics* 2. Oxford, OUP, 2012. Ebook

Further reading (optional)

- Portmore, Douglas. "Perform your best option." *The Journal of Philosophy* 110 (2013): 436–459. Link.
- Vessel, Jean-Paul. "Defending a Possibilist Insight in Consequentialist Thought." *Philosophical Studies* 142.2 (2009): 183-95. Link.
- Timmerman, Travis. "Does Scrupulous Securitism Stand-up to Scrutiny? Two Problems for Moral Securitism and How We Might Fix Them." *Philosophical Studies* 172.6 (2015): 1509-528. Link.
- Timmerman, Travis, and Yishai Cohen. "Moral Obligations: Actualist, Possibilist, or Hybridist?" *Australasian Journal of Philosophy* (2016): 1-15. Link.
- Vessel, Jean-Paul. "Against Securitism, the New Breed of Actualism in Consequentialist Thought." 28.2 (2016): 164-78. Link.
- Carlson, Erik. "Actualism and possibilism" in Carlson, Erik. *Consequentialism Reconsidered*. Dordrecht: Kluwer Academic, 1995. [not online] Link.
- Ando, 'Pluralism about reasons and agent-units in consequentialism'. Link
- Cohen, Yishai, & Travis Timmerman. "Actualism Has Control Issues." *Journal of Ethics and Social Philosophy*, 10.3 (2016): 1-19. Link.

Week 6: Global consequentialism (the inconsistency problem)

Core reading (mandatory)

- Adams, Robert Merrihew. "Motive Utilitarianism." *The Journal of Philosophy* 73.14 (1976): 467-81. Link.
- Feldman, Fred. "On The Consistency Of Act-Utilitarianism And Motive-Utilitarianism - A Reply To Adams, Robert." *Philosophical Studies* 70.2 (1993): 201-12. Link.
- Driver, J. "Global Utilitarianism" in Eggleston, Ben and Dale E. Miller. *The Cambridge Companion to Utilitarianism*. Cambridge, 2014: 166–76. Link.

Further reading (optional)

- Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon, 1984: 24-28. Ebook.

- Hooker, B. 'Rule consequentialism'. *Stanford Encyclopedia of Philosophy.* Stanford: Stanford University (1997).See especially the section on global consequentialism: Link.
- Railton, Peter. "How Thinking about Character and Utilitarianism Might Lead to Rethinking the Character of Utilitarianism." *Midwest Studies In Philosophy* 13.1 (1988): 398-416.Link.
- Kagan, S. "Evaluative focal points" in Hooker, Brad, Elinor. Mason, and Dale E. Miller. *Morality, Rules, and Consequences: A Critical Reader*. Edinburgh: Edinburgh UP, 2000. [not online] Link.
- Pettit, P. and M. Smith. "Global consequentialism" in Hooker, Brad, Elinor. Mason, and Dale E. Miller. *Morality, Rules, and Consequences: A Critical Reader*. Edinburgh: Edinburgh UP, 2000. [not online] Link.
- Crisp, Roger. "Utilitarianism and the Life of Virtue." *The Philosophical Quarterly (1950-)* 42.167 (1992): 139-60. Link.
- Streumer, Bart. "Can Consequentialism Cover Everything?" *Utilitas* 15.2 (2003): 237-47.Link.
- Lang, Gerald. "A Dilemma for Objective Act-Utilitarianism." *Politics, Philosophy & Economics* 3.2 (2004): 221-39. Link.
- Lazari-Radek, Katarzyna De, and Peter Singer. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford, 2014: 315-6. Ebook.
- Ord, Toby. *Beyond Action: Applying Consequentialism to Decision Making and Motivation* (2009). DPhil Thesis. [Available only to read in Weston Library]. Link.

Week 8: Emotion and rationality in ethics

Core reading (mandatory)

- Korsgaard, Christine M. "Skepticism about Practical Reason." *The Journal of Philosophy* 83.1 (1986): 5-25. Link.
- Slote, Michael. "Empathy: the cement of the moral universe" in Slote, Michael. *Moral Sentimentalism*. New York; Oxford: Oxford UP, 2009. Link.
- Maibom, Heidi. "What Experimental Evidence Shows Us about the Role of Emotions in Moral Judgement." *Philosophy Compass* 5.11 (2010): 999-1012. Link.

Further reading (optional)

- Kauppinen, Antti. "Sentimentalism" in LaFollette, Hugh. *The International Encyclopedia of Ethics*. Malden, MA: Wiley-Blackwell, 2013. Link.
- Gill, Michael B. "Moral Rationalism vs. Moral Sentimentalism: Is Morality More Like Math or Beauty?" *Philosophy Compass* 2.1 (2007): 16-30. Link.
- Todd, Cain. "Emotion and Value." *Philosophy Compass* 9.10 (2014): 702-12. Link.
- Slote, Michael. "Why not empathy?" in Cohen, I. Glenn, Norman Daniels, and Nir M. Eyal. *Identified versus Statistical Lives: An Interdisciplinary Perspective*. First ed. New York, 2015. Link.
- Haidt, Jonathan. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (2001): 814–34. Link.

- Greene, Joshua, R. Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen. "An FMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293.5537 (2001): 2105-8. Link.
- Greene, J.D. “The Secret Joke of Kant's Soul” in Sinnott-Armstrong, Walter, and Christian B. Miller. *Moral Psychology Volume 3*. Cambridge, Mass.; London, 2008: 35–80. [not online] Link.
- Singer, Peter. "Ethics and Intuitions." *The Journal of Ethics* 9.3 (2005): 331-52. Link.
- Kahane, Guy, Katja Wiech, Nicholas Shackel, Miguel Farias, Julian Savulescu, and Irene Tracey. "The Neural Basis of Intuitive and Counterintuitive Moral Judgment." *Social Cognitive and Affective Neuroscience* 7.4 (2012): 393-402. Link.

Reading lists for weeks 3,4,7.

Week 3: Moral Uncertainty

Core reading (mandatory)

- Bykvist 2017, “Moral uncertainty” 7pp – an opinionated introduction and guide to the literature
- Weatherson 2014, “Running risks morally” 23pp
- Sepielli 2016, “Moral uncertainty and fetishistic motivation” 18 pp

Further reading (optional)

- Harman 2015, “The Irrelevance of Moral Uncertainty” 26pp – another externalist argument, focusing on blameworthiness
- Mason 2015, “Moral Ignorance and Blameworthiness” 21pp – a nuanced take on different types of blameworthiness and obligation, a response to Harman and others
- Gustafsson and Torpman 2014, “In Defence of My Favourite Theory” 16pp – an internalist rival to expected choiceworthiness

Week 4: Moral Vagueness

Core reading (mandatory)

- Dunaway 2016, “Ethical Vagueness and Practical Reasoning” 23pp
- Schoenfield 2016, “Moral Vagueness is Ontic Vagueness” 26pp

Further reading (optional)

As Schoenfield mentions, part of the significance of her conclusion is widespread scepticism that there is such a thing as ontic vagueness. For an entry-point into this literature, see

- Barnes 2010, “Arguments Against Metaphysical Indeterminacy and Vagueness”, 11pp.

For slightly different takes on the metaethical implications of moral vagueness, see:

- Shafer-Landau 1995, “Vagueness, Borderline Cases, and Moral Realism” 14pp
- Constantinescu 2014, “Moral vagueness: a dilemma for non naturalism”, 34pp
- Dougherty 2014, “Vague value” 21pp

For another discussion of how to respond to vagueness, try

- Williams 2014, “Decision making under indeterminacy”, 68pp.

Week 7: Well-being: Internalism and Variabilism

Core reading (mandatory)

- Sarch 2011, “Internalism about a person's good: don't believe it” 24pp
- Lin forthcoming, “Welfare invariabilism” 30pp

Note: Sarch’s paper is a response to

- Rosati 1996, "Internalism and the good for a person" 31pp

Since Sarch provides a synopsis and quotes Rosati at length, my hope is that you won’t have to read Rosati’s paper separately (for now). But an obvious question when reading Sarch is whether you can help Rosati out, so it would be good to have her paper on hand.

Further reading (optional)

- Crisp 2017, “Well-being” (in *The Stanford Encyclopedia of Philosophy*) – look here if you don't know the basics
- Hall and Tiberius, “Well-being and subject dependence” 13pp (in *The Routledge Handbook of Philosophy of Well-Being*) – a survey of views friendly to internalism and variabilism
- Rosati 2006, “Personal good” (in *Metaethics After Moore*) – Rosati’s positive account of well-being.

Topics in Ethics – MT17 – Hilary Greaves and Teru Thomas

Week 1: Consequentialising

1. Consequentialising: the basic idea
    a. (Maximising) consequentialism: x is permissible iff x maximises the good.
        i. More formally: for all other available acts y, V(x) ≥ V(y), where the 'value function' V ordinally represents goodness of outcomes.
        ii. 'Outcome' here means a whole history of the world.
    b. Consider a theory that says that one ought not to break promises, 'even when the outcome would be better if one did'. This seems a paradigm example of a nonconsequentialist theory.
    c. Suggestion: But could it really just be a consequentialist theory that just happens to regard states of affairs in which promises have been broken as particularly bad?
    d. General worry (e.g. Nozick): This is 'gimmicky representation'. (I.e., it uses a notion of the good that is just cooked up to represent the given deontic facts, with no independent plausibility/rationale.) Isn't that giving up on the spirit of consequentialism?
2. Relativised value functions
    a. 'But my theory won't tell *me* to break a promise in order to prevent *you* from breaking two promises.'
        i. Agent-relativity: relativise the value function to agents. (It's right that I perform action X iff X maximises V_{me}.) Contrast Brown's 'agent-neutrality' axiom.
    b. 'But my theory also won't tell me to break a promise now in order to prevent *myself* from breaking two promises tomorrow.'
        i. Time-relativity: relativise the value function to times too ($V_{me,t}$).
    c. Other forms of relativity: relativity to choice set (to deal with choice-set-dependent theories; contrast Brown's "dominance" axiom) and/or to chosen action (to deal with theories that violate 'normative invariance')
3. Claim (the 'deontic equivalence thesis'): Every ('remotely plausible'?) moral theory can be consequentialised.
    a. Basic technique: x is permissible in circumstances C → assign the outcome of x an equal-highest value, relative to circumstances C.
    b. Problematic cases(?): supererogation, and moral dilemmas
        i. Moral dilemmas
            1. Problem: in a *moral dilemma* (by definition), all available acts are impermissible. But there will always be some available act that maximises value, if 'values' are real numbers.
            2. Portmore tries to accommodate moral dilemmas via normative variance. But a deontic theory might say "there are moral dilemmas and there is no normative variance".
            3. Suggestion: actually we need more exotic structures (for our value functions to take values in) than the real numbers, to accommodate genuine moral dilemmas.
        ii. Supererogation

        1. Problem: an act x is *supererogatory* iff (i) x is permissible and (ii) x is morally better than some other permissible act. But according to maximising consequentialism, the only permissible acts are those s.t. no other act is better.
        2. Portmore tries to accommodate this via his "dual-ranking act-consequentialism". (Does this move work?)
4. What's the *point* of consequentialising/observing that it can be done, insofar as it can be?
    a. Straw person 1 (?): Self-describing nonconsequentialists are confused; everyone is a consequentialist.
    b. ~Brown: Consequentialists are confused if they think they are advancing a non-trivial thesis; consequentialism has no content.
    c. Portmore, Dreier: By moving to the deontically equivalent consequentialist theory, the would-be non-consequentialist gets to hold on to the 'Compelling Idea' at the heart of utilitarianism/consequentialism, viz. that it is always permissible to bring about the most good.
        i. Schroeder: This doesn't work. Consequentialising gives us 'it's always permissible to bring about the good-relative-to-me-and-t-and-choice-set-C etc", but *that* wasn't the Compelling Idea. ("Good-relative-to…" is a new technical term introduced by the consequentialiser.)
    d. Dreier, Louise (?): We gain a natural taxonomy of theories by seeing *what it takes* to 'consequentialise' the theory (i.e. (substantively) which things we have to count as good, and (structurally) which forms of relativity we have to appeal to).
    e. Another suggestion: The consequentialised form of the theory is the one we need for dealing with (empirical or moral) uncertainty, using the existing tools of decision theory. So it's useful to see that (nearly?) all theories do have such a form.
        i. (Problematic theory-types for this purpose include: cyclic choiceworthiness orderings; massive incomparability)

Topics in Ethics – MT17 – Hilary Greaves and Teru Thomas

Week 2: Parity, incomparability, incommensurability

1. Orderings of alternatives
    a. Alternatives (outcomes, actions, etc.) are ranked by a relation 'at least as good as', written ≥. This relation is symmetric (everything is at least as good as itself), transitive (if x≥y and y≥z, then x≥z), *and complete* (for all x, y, either x≥y or y≥x (or both)).
    b. When a binary relation has these three properties, it is an *ordering*.
    c. We can define relations > ('strictly better than') and ~ ('equally as good as') from ≥:
        i. x>y iff (x≥y and not y≥x).
        ii. x~y iff (x≥y and y≥x).
    d. (Alternatively, *while we are assuming completeness of* ≥, we could have taken > as primitive and defined ~, ≥ from >.)
2. Incompleteness
    a. Sometimes it intuitively seems(?), of two items x and y, that neither is strictly better than the other, *but it is also not the case that the two are equally good*.
    b. Some examples
        i. Which of the options open to Sartre's student is better, staying at home to care for his mother or leaving to join the Resistance?
        ii. Which career is better for philosophical and musical Jane, that of a philosopher or that of a clarinettist?
        iii. Who is a better artist, Mozart or Michaelangelo?
    c. A minimal way of accommodating these intuitions: 'Incomparability'
        i. Deny that ≥ is complete. (Then ≥ is a 'preorder', rather than an ordering.)
        ii. Define > and ~ as before.
        iii. Also define: x¤y ("x and y are incomparable") iff (not x≥y and not y≥x).
        iv. For any alternatives x, y, either x>y, x~y, or x¤y.
    d. (Now we could not have defined ≥, ~ and ¤ using the single primitive >, because > does not enable us to distinguish between equal-goodness and incomparability.)
3. Some other standard options
    a. Epistemicism: One of the two options really is better than the other (or the two really are exactly equally good). It's just really hard to judge which.
    b. Indeterminacy: It's *indeterminate* which option is better than the other. (Employ e.g. supervaluationist machinery to model this, i.e. use a *class* of (complete) orderings.)
    c. 'Parity' (Chang):
        i. Postulate is an additional primitive value relation: 'on a par with'.
        ii. For any alternatives x, y, either x>y, x~y, x is on a par with y, or x¤y.
        iii. (In which case we need to revise some of our previous definitions: e.g. incomparability has to be redefined as (not x≥y, not y≥x *and* x and y are not on a par.)
        iv. Sartre's two options, Jane's two careers etc. are on a par with one another.
4. Some concerns about parity
    a. Chang says that parity, but not incomparability, enables choice between the alternatives to be 'justified', because 'justification presupposes comparability'. But why think this?

        i. What is clearly(?) right in ‘comparativism’: when choosing between A and B, what matters (for whether the choice of e.g. A is justified) is the relations between A and B, not the absolute features of A.
        ii. But in a context involving incomparability, we should presumably take ‘justification’ (=rational permissibility?) to require ‘maximisation rather than optimisation’ (cf. Hsieh).
    b. Why postulate an additional value relation? Unless the parity account is somehow superior to the other 3 existing accounts, Occam’s Razor tells against postulating it.
        i. Do Chang’s arguments for this superiority work?
    c. Do we really understand the difference between parity and incomparability?
5. Rabinowicz’s fitting-attitude apparatus
    a. Primitive notion: rational permissibility of preference
        i. For any set of alternatives X, let K be the set of rationally permissible (weak) preference relations on X. (Elements of K will presumably be reflexive and transitive, but need not be (complete) orderings.)
    b. Then we can define numerous ‘value relations’ in terms of rational permissibility, including (but not restricted to):
        i. x is better than y := it’s required to strictly prefer x to y (i.e., every element of K does this).
        ii. x is equally as good as y := it’s required to be indifferent between x and y.
        iii. x is on a par with y := it’s permissible to strictly prefer x to y, and permissible to strictly prefer y to x.
        iv. x and y are incomparable := it’s required to have a ‘preference gap’ between x and y.
    c. From this point of view, discussing exclusively ≥, >, ~, ‘on a par’ and ¤ looks arbitrary.
    d. ‘Choiceworthiness’ (= justified choice = rationally permissible choice?): a suggestion
        i. x is choiceworthy iff there is a permissible preference ordering in which no alternative is strictly preferred to x.
            1. (Why does Rabinowicz not consider this?)
        ii. Then there is no problem for “justified choice” under parity, but also none under incomparability.

Topics in Ethics – Teru Thomas and Hilary Greaves – week 2

Moral Uncertainty

Rough question: "How does what we ought to do depend on our beliefs/credences about norms?"

- What we ought to do *doesn't* depend on our beliefs about norms (~Weatherson and Harman).
- What we ought to do depends on our beliefs about norms *in pretty much the same way* it depends on our beliefs about empirical facts (~Sepielli).
- What we ought to do *does* depend on our beliefs about norms but *not* in the same way it depend on our beliefs about empirical facts (~Gustaffson-Torpman?).

Variations on this:

Suppose Jones does X rather than Y.

* Is it true that Jones ought to do X?

* Is Jones blameworthy for doing X? (Harman)

* Is it rational for Jones to do X?

Example 1. Jones thinks either total utilitarianism is true, or average utilitarianism is true, but she is not sure which. (In fact, total utilitarianism is true.) What should Jones do?

Jones ought to maximize total utility, but it might not be rational for her to do so.

Q. Is there any general principle connecting rationality and permissibility?

Types of Cases

(A) Jackson cases
(B) False beliefs < Especially emphasised by Harman
(C) Dominance cases < Main subject for Weatherson

(A) Jackson Cases

Since Weatherson and Sepielli don't discuss these cases, I'll stick to the

Standard empirical example. A doctor, having competently consulted all available evidence, is uncertain whether her patient has disease A or B: she takes both to be equally likely. If A, then pill X will cure and pill Y will kill; if B, then pill Y will cure and X will kill. Either way, pill Z will significantly relieve the symptoms. (In fact, the disease is A, so pill X will cure.)

Typical position: The doctor ought to prescribe Z, and would be rational and blameless to do so. It is impermissible to prescribe X, and the doctor would be irrational and blameworthy to do so.

* Nice consequentialist (or consequentialised?) story: maximize expected value.

* One way in which the doctor's obligations depend on her empirical uncertainty: if possible, the doctor is obliged to gather new empirical evidence. **Sepielli**: why are we obliged to improve our moral beliefs if not because of our normative uncertainty?

* “**Objectively**” the doctor ought to prescribe X, but “**subjectively**” she ought to prescribe Z.

> “Dr subjectively ought to Z” = “Dr ought to Z” (not: “Dr thinks she ought to Z”)
>
> “Dr objectively ought to X” = “Dr would be obliged to X if she knew the relevant facts” (?)

(See Broome, *Rationality Through Reasoning*, ch. 3, for some relevant discussion)

Q. What happens in “moral Jackson cases”?

(B) False beliefs

Weatherson’s case (taken from Harman):

> Hannah takes her spouse out for what is meant to be a pleasant anniversary dinner. It’s a nice restaurant, and there’s no reason to think anything will go wrong. But the restaurant gets bad supplies that day, and Hannah’s spouse gets very sick as a consequence of going there...
>
> Hannah should feel bad for her spouse, but there is no need for any kind of self-reproach.

Contrast:

> Hannibal is a 1950s father with sexist attitudes that were sadly typical. He has a son and a daughter, and makes sure to put together a good college savings fund for his son, but does not do the same for his daughter... As a consequence, his daughter cannot afford to go to college...
>
> Hannibal should feel ashamed, and guilty, about what he did. That’s because even if he had an excuse, **he did the wrong thing.**

Q. Is the difference here empirical vs normative, or justified vs unjustified belief?

- It's not clear whether Hannibal's false beliefs are normative rather than empirical.
- While Hannah’s beliefs are justified, it’s not clear that Hannibal’s are.

Q. Is there a cleaner pair of cases to help Weatherson here?

(C) Dominance cases

Weatherson’s main case:

> Martha is deciding whether to have steak or tofu for dinner. She prefers steak, but knows there are ethical questions around meat-eating. She has studied the relevant biological and philosophical literature, and concluded that it is not wrong to eat steak. But she is not completely certain of this; as with any other philosophical conclusion, she has doubts. As a matter of fact, Martha is right in the sense that a fully informed person in her position would know that meat-eating was permissible, but Martha can’t be certain of this. What should she do?

Weatherson: it *is* permissible for Martha to eat the steak.

Uncertaintists: it *isn’t*; Martha ought to avoid eating meat.

Weatherson on the 'Might' Argument

Weatherson thinks uncertaintists must be thinking along the following lines:

1. In the circumstances that Martha is in, eating a steak might be morally wrong.
2. In the circumstances that Martha is in, eating vegetables is definitely morally permissible.
3. Missing Premise
4. So Martha should not eat the steak.

Simplest version of the missing premiss:

> **ProbWrong.** If an agent has a choice between two options, and one might be wrong, while the other is definitely permissible, then it is wrong to choose the first option.

He then argues that ProbWrong is incorrect: if Martha knew ProbWrong, she would be rationally compelled to believe that eating steak is wrong; but actually her situation of uncertainty is perfectly rational.

Q. Is Weatherson's analysis of ProbWrong correct? Do ProbWrong and the Might Argument capture the thought behind uncertaintism?

> **Cake.** Carla is baking a cake for a fundraiser. She wants to put some sweetening syrup into the cake to improve its taste. She reaches for an unmarked bottle, which she is pretty sure contains the sweetener she wants. But then she remembers that last week she had some arsenic in a similar bottle. She is pretty sure she threw the arsenic out, but not exactly certain. As a matter of fact, the syrup in the bottle is sweetener, not arsenic, but Carla isn't certain of this. What should she do?

Q. What's the argument that Carla should not use the bottle?

> **? ProbObjectivelyWrong.** If an agent has a choice between two options, and one might be *objectively* wrong, while the other is definitely *objectively* permissible, then it is [subjectively] wrong to choose the first option.

Weatherson's Preferred Analogy

Recall. Bob has some credence that going to the gallery would be good for him (but it wouldn't be).

- Bob's welfare is such that it is irrational [i.e. imprudent?] for him to do something that might undermine it for no compensating gain.
- NOT: It is irrational for Bob to do something that might undermine his welfare, whatever that turns out to be, for no compensating gain.

So Bob would not (as far as this goes) be "irrational" to avoid the gallery.

By analogy,

- Morality is such that it is wrong for Martha to do something that might go against it;

- NOT: it is wrong for Martha to do something that might go against morality, whatever that turns out to be.

So it would not (as far as this goes) be wrong for her to eat the steak.

Q. Is Weatherson right about Bob? Is the analogy a good one?

Weatherson and Smith on “Fetishism”

Smith exegesis

> **Smith 1994:** Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read de dicto and not de re. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue.

Ice Cream: I like ice cream, and I judge that giving money to the shop is the way to get ice cream. Then I'm motivated to give money to the shop. But I don't really *care* about giving money to the shop; it’s just the means to an end. My motivation to hand over the money is "derivative" of my standing desire for ice cream.

“Fetishistic” motivation: I want to do what is right; I believe that helping cows is right; therefore I am motivated to help cows. But I don’t really *care* about helping cows; it’s just a means to the end of doing what’s right.

So to get the right motivational pattern, Smith thinks we should endorse something like

Motivational Internalism. Rationality requires that, if you believe that helping cows is right, then you care (non-derivatively) about helping cows.

Then the connection between belief and motivation is not mediated by a standing desire to do what's right.

Possible Weatherson: Motivational Internalism can’t explain why Martha would be motivated not to eat steak; she *doesn't* believe that helping cows is right, and, more generally, she *doesn't* value cows; she just thinks that it *might* be appropriate to value them. If she were to avoid eating meat, her motivation would only be derived from a "fetishistic" standing desire to do what's right. This *isn't* how moral motivation works in virtuous people, so… Martha wouldn’t be responding virtuously to her predicament if she avoided eating meat; so… it can’t be that she ought to avoid eating meat.

Sepielli doubts every step of the argument:

- Is it really true that Martha would have to have “fetishistic” motivation?
- Is this kind of motivation really bad in cases of uncertainty, if the alternative is a "leap of faith"?
- No obvious connection between questions of motivation and whether an action is “in conformity with duty.”

Q. Any help for Weatherson here?

Topics in Ethics – Teru Thomas and Hilary Greaves – Week 4

Moral Vagueness

INTRODUCTION

Main idea: different theories of vagueness have different implications both for first-order ethics and for meta-ethics

(1) **What is vagueness?** Hard to say anything completely theory-neutral.

- Indicated by a *sorites series*, a series of cases $C_1,...,C_{100}$ and a predicate "P" such that $P(C_1)$, not $P(C_{100})$, and yet for each i it's unassertable that $P(C_i)$ and not $P(C_{i+1})$.
- Associated with the existence of borderline (indefinite, indeterminate, unclear) cases – cases C for which it's unassertable that P(C) and unassertable that not P(C).

(2) **Theories of vagueness** – vary in accounts of *logic*, *type*, and *cognitive load.*

Logic: most obviously an issue of bivalence, but also questions about disquotation, modal logic, and much else.

Type:

"Semantic" – the referent of "P" is "undecided" between various precise properties (precisifications, sharpenings). Typically: It's true that P(c) just in case c has *all* of these properties, and it's false that P(c) just in case c has *none* of them. A borderline case has *some* of these properties but not others.

"Epistemic" – the referent of "P" is some specific precise property, but it's practically impossible to know which one. Williamson: if Sam is 180cm, then "Sam is tall" and "Sam is not tall" could each easily have been true. But knowledge requires safety.

"Ontic" – anything else (according to Barnes)

Cognitive Load: What kinds of attitudes should we have towards borderline cases?

NOTES ON DUNAWAY

Question: What "rationally ought" we to do given that we know it's indeterminate what is morally permitted? Modest answer: it depends on your theory of vagueness.

Note. He is primarily interested in consequentialism, so moral vagueness boils down to value vagueness. He talks mainly about vagueness of permissibility, but this might be a mistake?

Simple options:

* X is borderline better than Y >> X is (at least rationally) permissible (like parity – Broome)

* X is borderline better than Y >> borderline whether rational to X

Dunaway claims *epistemicists* and perhaps some others are bound to maximize expected value, but what this amounts to depends on the theory.

Standard picture: one assigns each possible world a *credence* between 0 and 1 and a *value;* it's rational to choose an act with maximal *expected value*. But if the value function is indeterminate, it's indeterminate what it's rational to do.

Dunaway: given epistemicism, we also have credences over *precisifications of the value function*; if we average with respect to these we get a determinate expected value.

Example: Sorites series C_n of creatures from definitely not sentient to definitely sentient. Options: maim creature C_n or kill one human. Killing the human costs 100 utiles. If the C_n are sentient, maiming them costs 1000 utiles.

The main question: what do appropriate credences over precisifications look like?

- Credence that C_n is sentient increases steadily from 0 to 1 as n increases. This looks natural for epistemicists and maybe some others.
- (Field) Credence that C_n are sentient, and credence they are not sentient, are both 0 or close to it in borderline cases >> breakdown of probability theory, potentially a lot of permissibility.
- (N. Smith) Credence that C_n are sentient tracks the "degree of truth" that they are sentient. Like 1, but don't satisfy probability theory.

He also considers another "degree of truth" picture (Degree-Value Connection): it is rationally permissible to choose X iff the degree to which X is wrong is minimal.

NOTES ON SCHOENFIELD

Central thesis: If a robust form of moral realism is true, and there is moral vagueness, then it is ontic vagueness.

Definitions:

"moral vagueness": primarily concerned with vagueness of permissibility

"robust moral realism": "moral truths are necessary and… moral properties [and in particular permissibility?] 'are part of the deep underlying metaphysical structure of the world, and they obtain entirely independently of how we conceptualize the world'" [This sounds very strong.]

"ontic vagueness": "Vagueness in how things are… Vagueness that would remain even if we spoke a perfect language and were omniscient."

"Perfect language": "a language ["ontologese"] that contains all and only predicates that are necessary to provide a complete and accurate description of how things are fundamentally."

[*Naive argument*: a perfect language must have a word – say, "guff" – for permissibility. But permissibility has borderline cases, so guff would also have borderline cases.]

AGAINST SEMANTIC ACCOUNTS

To be denied: In a perfect language moral predicates would exist [<< moral realism], but they would have precise application conditions [<< no ontic vagueness] even though they don't in English [<< semantic vagueness].

Basic argument: semantics of "permissible" isn't shifty, so it's rigid; if rigid, then vagueness is ontic.

> **The shifty view:** The truth-value of an utterance: "J is permissible" is highly sensitive to the way the word "permissible" is used in a ["liberal"] linguistic community. A sentence "J is permissible" may have one truth-value when uttered by S, in English, but a different truth-value when uttered by S', who is in all respects just like S, except that S' lives in a linguistic community that applies the predicate "permissible" slightly more liberally than we do.
>
> **The rigid view:** The shifty view is false. Slight changes in the application of "permissible" will not affect the truth-values of sentences containing this predicate. (p. 264)

Why the semantics of permissibility isn't shifty

> **The Argument:** "What does… follow from the shifty semantic account is that Cheryl can find out that some abortion [at 150 days], whose permissibility she was uncertain about, is, in fact (determinately!) permissible by collecting linguistic data… However, it doesn't seem like crunching through linguistic data is a way of resolving doubts about the permissibility of abortion".

Another try: No reason to care about permissibility rather than whatever property S' picks out with "permissible" ??

The problem with rigid views

1. Reference Magnetism

The argument:

(a) if there's a unique most natural precise property in the vicinity, then "permissible" must refer (precisely) to it; Boo!; so there must be a range of equally natural precise properties in the vicinity.

(b) if variations in usage can vary the range of semantic indecision among these most natural precise properties, then permissibility is shifty; Boo!; so variations in usage can't vary the range of indecision.

(c) if variations in usage can't vary the range of indecision, then "permissible" even as used by speakers of a perfect language, would be undecided over the same range of precise properties; so there is ontic vagueness.

2. **Conceptual Roles**

Focus on betterness rather than permissibility. Wedgwood: it is characterised by conceptual role of "making valid" inference from "X is better than Y" to a preference for X over Y.

If there's vagueness, is it because

(I) It's indeterminate which conceptual role is characteristic? – i.e. which inference rules?

(II) It's indeterminate "which relation makes those rules valid"?

On (I): Resolving this indeterminacy would determine cutoffs for "better than".

Note. Schoenfield is thinking of the candidate conceptual roles as, for example, the roles of "subjective" vs "objective" betterness. Are there more relevant ways in which the conceptual role might be indeterminate?

On (II): if filling a certain conceptual role is what's important about betterness, then a perfect language must have a name for whatever fills that role; if it's indeterminate, then we have ontic vagueness.

AGAINST EPISTEMIC ACCOUNTS

Epistemicists need to explain where the ignorance comes from and why it's a special form of ignorance (they are trying to give an account of vagueness). The claim here is that no known account gives an explanation that is plausible in the case of permissibility.

Semantic plasticity. *Williamson* explains the ignorance in terms of shifty semantics.

Truthmaker gaps. *Sorensen* suggests that in borderline cases there are no truthmakers, and we can't know borderline truths because we aren't acquainted with their truthmakers. Response: "truthmakers were never part of the moral realist's moral epistemology" Also: on Sorensen's view, the relevant moral facts are unknowable even in principle.

Arbitrary reference. *Kearns-Magidor*: reference of vague predicates is "arbitrary". This seems worse than shiftiness – who cares about the outcome of a reference lottery?

Topics in Ethics – MT17 – Hilary Greaves and Teru Thomas

Week 5: Actualism, possibilism and securitism

1. The central example: Professor Procrastinate
    a. At t, Procrastinate receives an invitation to review a book. It would be best if he wrote the review. However, he is a terrible procrastinator; if he were to accept the review assignment, he would in fact never get around to writing the review, and would just waste everybody's time. In order from best to worst, the possibilities that are in some sense open to him are: accept and write the review (best), decline, accept but then fail to write the review (worst). Procrastinate knows all this. Ought he to accept or decline the review invitation?
    b. *Actualists* hold that Procrastinate ought to decline (on the grounds that accepting would in fact lead to worse consequences than declining).
    c. *Possibilists* hold that Procrastinate ought to accept (on the grounds that he ought to accept and then write, since this is the best option).
2. Actualism: Some rival precise formulations
    a. For any action A that is possible for S, A ought to perform A iff the result would ensue if S did A is better than the result that would ensue if S did not do A.
        i. A ought to decline.
        ii. A ought to accept-and-then-write.
    b. Relative to any choice set $\{A_1, \ldots, A_n\}$ of actions, S ought to perform A_i iff the result that A_i would lead to is better than the result that any other action in the set A would lead to.
        i. Relative to the choice set {accept, decline}, A ought to decline.
        ii. Relative to the choice set {accept-and-then-write, accept-and-then-not-write, decline}, A ought to accept-and-then-write.
3. Possibilism: Precise formulation
    a. S ought to perform A iff A is part of the most choiceworthy maximally specific course of action that is performable for S.
        i. X is 'performable' for S: There is some schedule of intentions I such that S would X if S had I over the relevant time period.
        ii. Procrastinate ought to accept-and-then-write.
        iii. Procrastinate ought to accept.
4. Objections to actualism
    a. Objection 1: Actualism is inconsistent, since it says that Procrastinate ought to accept-and-then-write, but denies that Procrastinate ought to accept.
    b. Reply: Not *inconsistent*. Rather, actualism violates
        (**ODC**: Obligation distributes over conjunction): O(A&B) → [O(A) & O(B)].
        "If one ought to do A-and-B, then one ought to do A and one ought to do B."
        (Consider A=accept, B=write.)
        i. Actualists are usually quite happy with this violation. (For dissent, see Vessel, 'Defending a possibilist insight in consequentialist thought'.)
    c. Objection 2: According to actualism it is impossible for Procrastinate to fulfil all of his obligations. (He cannot both decline, and accept-and-then-write, but according to actualism he has obligations to do each of these.)

d. Reply: it's possible for Procrastinate to be such that he meets all of his obligations (since it's possible for him not to be a procrastinator). What is impossible is only fulfilling all of the obligations that Procrastinate *in fact* has, given his imperfections.
    i. Arguably a *good* feature of the actualist theory: It is intuitively the case that Procrastinate (if he declines) meets some but not all of his obligations.
e. Objection 3: Actualism makes it too easy for an agent to avoid incurring moral obligations, simply by being sufficiently morally imperfect.
    i. E.g. Procrastinate avoids having an obligation to accept the review assignment.
    ii. Reply: Procrastinate does not avoid having an obligation to accept-and-then-write.
f. Objection 4: Actualism judges that agents are permitted (and even required) to do terrible things, as long as what that agent would in fact do otherwise is even worse.
    i. Ross's example: Ace ought to give Emily Drano (because although he *could* give her water, what he would *in fact* otherwise do is give her arsenic).
g. Objection 5: Actualism generates too many obligations.
    i. Ross's example: Ace ought to give Emily Drano, and Ace ought to give Emily bleach, and…
    ii. Reply: This follows only from the first formulation of actualism (above), not from the second. (But maybe the first formulation is superior?)

5. Objections to possibilism
    a. Objection 1: Procrastinate ought to decline!
        i. From the point of view of Procrastinate at t, there's no relevant distinction between Procrastinate's own future actions and the actions of someone else. (Clearly Procrastinate ought not to accept *on behalf of his colleague Lazy* if he knows that Lazy would not actually write the review.)
    b. Objection 2: Possibilism is arbitrary
        i. It's true that saying yes is part of the best possible course of action open to Procrastinate, but it's also part of the worst course of action. Why break the symmetry in one direction rather than the other? (Note that actualism, but not possibilism, has an answer to this.)
    c. Objection 3: Possibilism is in tension with obvious facts about moral advice
        i. Clearly a third party ought to *advise* Procrastinate to decline. How come then that Procrastinate himself should not decline?
    b. Objection 4: Possbilism requires denying 'detachment for conditional obligation'
        i. (DCO): If I ought to do X given Y, and Y is the case, then I ought to do X.
        ii. Consider X=decline, Y=Procrastinate would not write if he accepted.
6. Securitism (see esp. Portmore, *Commonsense consequentialism: Wherein morality meets rationality*, chapter 6, section 3)
    a. Basic idea: The reason Procrastinate ought (as of t) to decline is that *there is nothing he can now do to 'secure'* the state of affairs in which he accepts and writes, rather than accepts and fails to write. ('Secure': via his intentions and other attitudes at t.)
    b. Securitism, roughly: What an agent ought to do (as of t) is:
        i. the best course of action that is 'securable' by her at t;
        ii. anything that is part of that best course of action.
    c. Application to the case of Procrastinate:
        i. The courses of action that are securable by Procrastinate at the time of decision are {decline, accept-and-fail-to-write}.

    ii. The best of these is *decline*.
    iii. Therefore Procrastinate ought to decline.
d. Objection: accept-and-later-write *is* securable by Procrastinate at t. E.g., Procrastinate could secure that option by adopting a belief that his life depends on writing the review (thus ensuring that he later has sufficient motivation to write).
e. 'Scrupulous' securitism: rule this out, by adding a clause to the effect that the 'securing' mechanism must not involve any *impermissible* attitudes.
f. Objection/question: Suppose that Procrastinate could secure accept-and-later-write by now taking a pill that makes him later believe that his life depends on writing the review. *Intending to take the pill* is not irrational. So does this make accept-and-write scrupulously securable?
    i. Yes. But this doesn't seem to be a problem. Plausibly in this case Procrastinate *should* accept.

7. Ross's objection to (simple and scrupulous) securitism
    a. Background observation: What is [scrupulously] securable by a given agent S varies over time. Therefore a securitist theory will have to relativise obligations to times.
    b. However, the *content* of an obligation held at t can include the agent's doing certain things at *other* times t'≠t.
    c. Example: Procrastinate has an obligation *at t* that he *later* write the review, *if* his later writing the review is [scrupulously] securable by Procrastinate at t.
    d. Ross thinks that on an adequate theory, one must be obligated *at t* to do the best one can *at t'*, even if the best one can do at t' is not something that is securable by the agent at t. He then objects that securitism does not deliver this (e.g. in "Satan's school for girls").
        i. It's not clear this requirement is legitimate, though.
8. Temporal vs atemporal examples
    a. The example of Procrastinate is *temporal*, in the sense that Procrastinate's problematic deficiency is due to manifest itself at a time *other than* the time of decision we are primarily interested in.
    b. An atemporal analog (Jackson and Pargetter): Jones is driving in a tunnel, where she should not change lanes, but she is going to change lanes. If she changes lanes, then accelerating is better (it disrupts the flow of traffic less). If she doesn't change lanes, then accelerating would lead her to drive into the back of a truck. The betterness ordering of outcomes is (don't change and don't accelerate, change and accelerate, change and don't accelerate, don't change and accelerate). Should Jones accelerate?
        i. Actualists: Yes.
        ii. Possibilists: No, because the best performable course of action is [don't change lanes and don't accelerate], and don't-accelerate is part of this.
        iii. Securitists: No, because the compound action [don't change lanes and don't accelerate] is securable by Jones at the time in question.
            1. Unless we can somehow 'silo off' Jones' fixed tendency to change lanes, so that securitists agree (with actualists) that this tendency should be held fixed. But it's not clear whether or how this could be done.
            2. This is a bit odd. Securitism was supposed to basically side with actualism.

Topics in Ethics – MT17 – Hilary Greaves and Teru Thomas

Week 6: Global consequentialism and the inconsistency problem

1) Background: Consequentialism and the self-defeatingness worry
   a) Act-consequentialism
      i) Act-consequentialism (AC), roughly: An act is right iff no other available action has higher [expected] value.
      ii) Two criticisms
         (1) (AC) fails to issue verdicts on anything other than acts. But motivations, character traits etc are also important moral evaluands.
         (2) (AC) *suggests* a way of evaluating motivations etc that (however) is in tension with (AC) itself… [TBC]
   b) Decision procedures
      i) Two natural ways to extend AC to decision procedures
         (1) (DC1) Agents ought to decide what to do by considering which available action has higher [expected] value.
         (2) (DC2) Agents ought to decide what to do using whichever decision procedure has the property that following that decision procedure leads to higher [expected] value than any other available decision procedure.
      ii) These 'criteria of right decision procedure' can come apart from one another, for at least three reasons
         (1) Calculation time: The process of considering which available action has higher [expected] value is sometimes too time-consuming to be worth the cost (e.g. when deciding whether or not to save someone from drowning).
         (2) Personal bias: Agents who try to work out which acts are recommended by AC 'in the heat of the moment' might have a systematic tendency to skew their estimates in the direction of their own self-interest.
         (3) 'Alienation': Going through the process of the AC calculations might 'alienate' the agent from valuable natural emotions and motivations, and thereby substantially reduce the agent's quality of life.
      iii) The 'self-defeat' worry: "using (AC) as decision procedure" (i.e., conforming to (DC1)) would make one less likely to act rightly by the lights of (AC) itself.
      iv) Standard response ('two-level consequentialism'): Accept (AC)&(DC2), not (AC)&(DC1).
      v) Sort-of-criticism (Williams): This is tantamount to giving up (AC)
         (1) "There is no distinctive place for direct [i.e. act] utilitarianism unless it is, within fairly narrow limits, a doctrine about how one should decide what to do. This is because its distinctive doctrine is about what acts are right, and, especially for utilitarians, the only distinctive interest or point of the question what acts are right, relates to the situation of deciding to do them."
         (2) Note though that even if Williams is right about this, accepting (DC2) is not tantamount to giving up *consequentialism*.
   c) Motivations
      i) Similarly, AC can be extended to evaluate motivations in two fairly natural ways
         (MC1) Agents' motivation ought to be maximisation of [expected] value.
         (MC2) Agents' motivations ought to be whatever pattern of motivations has the property that having those motivations has higher [expected] value than does having any alternative pattern of motivations.
      ii) These 'criteria of right motivations' can similarly come apart from one another

(1) Parfit's example of Clare: "Clare could either give her child some benefit, or give much greater benefits to some unfortunate stranger. Because she loves her child, she benefits him rather than the stranger."
    (a) According to AC, benefitting the stranger is the right act.
    (b) Benefitting the stranger is also what Clare would do if she had the pattern of motivations prescribed by (MC1).
    (c) According to (MC2), though, the right pattern of motivations for Clare includes love for her child, although Clare *would not* benefit the stranger if she has this motivation.
        (i) (*Could* Clare benefit the stranger while loving her child? This is not clear.)
(2) Adams' example of Jack (at the cathedral in Chartres): This case has the same relevant structure as that of Clare.

d) Generalising: Global consequentialism
    i) Pettit and Smith: the right X, for any evaluative focal point X, is the best X, i.e. the one that maximises [actual or expected] value.
        (1) 'Evaluative focal points': things that are appropriate objects of moral evaluation.
        (2) Includes: motivations, character traits, intentions, decision procedures, maybe public institutions… (rather than just acts).
        (3) Possibly excludes: climates, eye colours, …
    ii) Problem/unclarity: in general it's not clear how to assign values to evaluands, since evaluands are not propositions. E.g. the value of *having* the character trait of generosity might be quite different to the value of *trying to inculcate* the value of generosity.
    iii) Fix: relativise to 'role': "for any evaluative focal point X and role R, the right x∈X in R is the one such that having x in R has higher [actual or expected] value than having any other y∈X in R."
e) Distinguish: Indirect vs global consequentialism
    i) Indirect consequentialism
        (1) Evaluate *one* 'evaluative focal point' via the consequentialist formula 'the right X is whichever X maximises [expected value]'.
        (2) Evaluate every other focal point derivatively (or not at all).
        (3) Examples
            (a) Rule-consequentialism: direct consequentialist evaluation of rules; indirect derivative evaluation of acts, via the criterion of conformity to the right set of rules
            (b) Motive utilitarianism: direct consequentialist evaluation of motives; right action as whatever does/would issue from right motives
    ii) Global consequentialism
        (1) Evaluate every 'evaluative focal point' *directly,* via the consequentialist formula (as above).
    iii) Pettit and Smith: No form of indirect consequentialism is plausible. Global consequentialism is the only form of consequentialism that might be correct.
f) The inconsistency problem for global consequentialism
    i) Basic thought: There seems to be some tension in accepting e.g. (AC)&(MC2), since e.g. it leads us to give both a positive and a negative evaluation of Clare (*whatever she does*, provided only that what she does is consistent with the 'laws of psychology').
    ii) Note that the tension is not literally a *contradiction*. (No commitment to P&~P, for any P.) What then is the nature of the problem, exactly?

(1) Adams: [(AC)&MC2)] are "incompatible in the sense that holding the latter ought reasonably to prevent us from holding the former as a *moral* theory", since (MC2) entails that "it is morally better on many occasions to be so motivated that one will not even *try* to do what one ought, by [(AC)] standards, to do". (Not clear how this is supposed to follow.)
(2) Hooker: GC gives incoherent guidance. In a case involving conflict between act and decision procedure, "global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure. That seems paradoxical."

g) Feldman's reply
   i) The appearance of 'inconsistency' is an artefact of formulating consequentialism inadequately, not of trying to combine (e.g.) AC with MC2 specifically.
      (1) Can already arise under AC alone: e.g. suppose I could take a cold remedy, so that I have a better day overall, but would as a result make some suboptimal decisions later (due to drowsiness). In this case, according to AC taking the cold remedy is right, but making the suboptimal later decisions is not.
   ii) Feldman's formulation
      (1) For a given time t and person S, certain worlds are 'accessible' to S at t
      (2) If w is accessible to S at t, and no better world is also accessible to S at t, say that w is *a best* for S at t
      (3) "U: As of t, for all propositions p, S morally ought to see to the occurrence of p iff p is true in all the bests for S at t"
      (4) Special cases: p could be a proposition whose content is that *S performs act x*, or that *S has motivations m*
      (5) Consistency guarantee: Feldman's theory U will only ever say that *S ought to see to S's doing act x* and also that *S ought to see to S's having motives m* if doing x and having m are compossible. (The key decider is which actions S performs and which motives S has in *the best* world that is accessible to S at t.)
      (6) 'Accessible' here *could* mean e.g. 'performable' or 'securable' (cf. week 5)
   iii) Implications of Feldman's theory for the case of Clare
      (1) Key question: At t0, is there an accessible possible world in which Clare loves her child and yet benefits the stranger?
      (2) If yes, then:
         (a) At t0, Clare ought to see to it that she loves her child.
         (b) At t0, Clare ought to see to it that she benefits the stranger (e.g. by forming the intention to do so).
         (c) By t1, if Clare has done as she ought, she loves her child.
         (d) At t1, the world in which Clare loves her child and yet benefits the stranger is (presumably) still accessible.
         (e) At t1, Clare ought to benefit the stranger.
      (3) If no, then:
         (a) At t0, there are accessible possible worlds in which Clare loves her child, and there are accessible possible worlds in which Clare benefits the stranger, but there is no accessible possible world in which both of these conditions hold.
         (b) At t0, Clare ought to see to it that she loves her child.
         (c) At t1, if she has done as she ought, Clare loves her child. As a result, possible worlds in which Clare benefits the stranger are no longer accessible.

(d) At t1, Clare ought to benefit her child (since there is no better accessible alternative).

(4) Towards *resolving* the key question

(a) Presumably it is true at t0 that: if at t1 she loved her child, then Clare *would not* benefit the stranger. But on the standard semantics for counterfactuals, this is just a matter of who Clare benefits at the *closest* possible world in which she loves her child.

iv) Resisting Feldman's model

(1) Not clear why the key thing should be whether p is true in all the *best* accessible worlds. Why not instead just consider whether p is better ("on average") than the alternatives?

h) Driver's (different) reply

i) 'Normative ambivalence': In cases in which the tension arises, it is *appropriate* to have both a positive and a negative evaluation of the agent (e.g. a positive evaluation of the agent's motives, and a negative evaluation of her act).

ii) "It is not at all surprising that guidance might be mixed, because guidance about what to *do* is different from guidance about what to *be*." (How satisfying is this?)

Topics in Ethics – Teru Thomas and Hilary Greaves – Week 7

Well-being: internalism and variabilism

I. Rosati on Internalism

Platitude: If X is good for S, then X is desirable for S's sake.

Extreme Anti-Paternalism: If X is good for S, then S actually desires X (or has some other pro-attitude).

Simple Internalism: ...S herself could desire X

Strong Internalism: … in ideal circumstances, S would desire X for her actual self.

According to Rosati, we must explain "ideal circumstances" in a way that respects S's actual nature and in particular her capacity for "rational self-governance"

Two-Tier Internalism (simplified a little).

1. Strong Internalism
2. Def. "Ideal circumstances": S herself would ordinarily regard her attitude towards X in those circumstances as authoritative.

Sarch's reformulation:

> If X is good for S then S has a "**counterfactually sanctioned desire**" (maybe not an actual desire) for X.

ROSATI'S ARGUMENTS/MOTIVATIONS FOR INTERNALISM

1. Argument from Judgement Internalism: S's judgement that X is good for S ordinarily induces S to care about X. But this is mysterious if for some X that is good for S, S *can't* care about X. (310ff)

2. Argument from Metaphysics: The motivational aspect of goodness is metaphysically mysterious unless "the very goodness of her good is *constituted* by her being disposed to care about it." (313ff)

3a. Argument from Epistemology: we could only *justifiably believe* that X is good for S if S is capable of caring about X; that indicates something about the nature of goodness-for-S (315ff)

3b. Argument from Reasons Internalism: S can only have a reason to promote X if S has, or could have, some proattitude towards X; but S *does* have a reason to promote her good. (318ff)

4. Argument from "Ought Implies Can": S *ought* to desire her good, so she *can* desire her good. (320ff)

5. Argument from Autonomy: “something can suit or fit a person and hence be good for her only if, in some sense, she could come to care about it on her own.” (322ff)

II. Example: Object CS-Desire Satisfaction

Generic desire-satisfaction view: What’s good for S is what S desires.

Basic problem: People frequently desire things that are bad for them.

More plausible: invoke S’s rational, counterfactually sanctioned [cs], or otherwise “cleaned up” desires rather than S’s actual desires. I’ll be loose about this.

EDS *Episode* Desire Satisfaction: What’s good for S are *episodes of desire satisfaction*. (Equivalent: What’s good for S is “what S desires” de dicto.)

ODS *Object* Desire Satisfaction: What’s good for S are *things S desires.* (What’s good for S is “what S desires” de re.)

> Note: “objects” of desire naturally understood as *properties*

Clarification: Two pictures of what a theory of good-for-S should do.

- Whether, in w0, w is better for S than v, depends only on what properties S has in w and v. So, if true, it’s *necessarily* true that w is better for S than v: there is a single ranking of worlds. **A theory of good-for-S should specify a ranking of worlds.** (The default picture; EDS, not ODS, is compatible with this)
- Whether, in w0, w is better for S than v, may also depend on what properties S has in w0. So it’s contingent whether w is better-for-S than v; different rankings obtain in different worlds. **A theory of good-for-S should specify a function from worlds to rankings of worlds.** (ODS is compatible with this)

Sarch: Of ODS and EDS, (i) only ODS satisfies internalism; (ii) only EDS is plausible.

> On (i): Internalism says: if X is good for S, X is the object of a cs-desire. ODS makes this true. But no reason to think that *episodes of cs-desire satisfaction* must be objects of cs-desire. For similar reasons, hedonism, etc., seem not to be internalist.
>
> On (ii): Sarch mentions some theoretical obstacles for ODS. But EDS has prima facie problems of its own:

Against EDS: On EDS, a prudent strategy for S would be to obtain *lots* of desires for what is likely to be true.

Possible Take-Away: Put *some* weight on Rosati’s arguments and develop something like ODS.

III. Lin on Variabilism

Puzzle: Is internalism/ODS compatible with Lin's arguments?

(A) First formulation: "Invariabilism is the view that the same theory of welfare is true of every welfare subject. Variabilism is the view that invariabilism is false."

> **ODS is compatible with invariabilism, by this standard.** Cf his discussion of "perfectionist" and "sophisticated" views.

(B) Second formulation: "Variabilism implies either that no list of basic goods and bads applies to all subjects or that the basic prudential values of the tokens of at least one basic good or bad are calculated differently for different subjects"

> **Ambiguous whether ODS entails variabilism.** In general, internalism suggests variabilism insofar as different subjects have different cs-desires.
>
> **E.g. Lin on desire-satisfaction:** "it claims that the kind *thing that satisfies one of your desires* is the only basic good."
>
> But on ODS this phrase picks out different kinds for different people.
>
> > Ann desires company; *having company* is of the kind "thing that satisfies one of Ann's desires"
> > Bob desires solitude; *lacking company* is of the kind "thing that satisfies one of Bob's desires"; *having company* is not.
>
> So it's not that the same kinds are basic goods for different people.
> Contrast: "Episode of desire satisfaction" picks out the same kind for different people.

Lin's characterisation of **basic goods** might seem to rule out ODS:

(C) "the claim that K is a basic good for a certain subject, S, merely implies the following about the possession by S of tokens of K: **for any tokens of K, if S were to have them, each of them would be basically good for S**"

> [Notes: (1) He seems to think this is also a *sufficient* condition for K to be a basic good. (2) It helps him if the counterfactual is sometimes false even when the antecedent is impossible. This is controversial (see e.g. Williamson, *The Philosophy of Philosophy*).]

Example. Suppose that if Contrary Cathy were to have company, she would desire solitude, and vice versa. She actually desires company.

> ODS: having company is basically good for Cathy; if Cathy had company, having company would not be basically good for Cathy.
>
> (C): If having company is basically good for Cathy, then, if Cathy had company, that would be basically good for Cathy.

But the counterfactual in (C) is ambiguous:

> Since having company is good for Cathy...
>
> (C1) if Cathy had company, that would be basically good for her according to the ranking of worlds that would *then* obtain?
>
> (C2) if Cathy had company, that would be basically good for her according to the ranking of worlds that *actually* obtains?
>
> ODS as stated is compatible with C2, not C1. But only C2 seems to be a plausible account of what it means for having company to *actually* be good for S.

Example from Lin, on perfectionism:

> Why deny, then, that theoretical contemplation is a basic good for Fido if it is a basic good for us? Why insist that Fido's actual nature determines what he would benefit from in far-out counterfactual scenarios in which he has a very different nature?

It's ambiguous what the dubious claim is meant to be.

Possible Take-Away: Lin does not make a clear argument against variabilism in the sense of ODS. But it isn't easy to talk in a way that makes ODS sound plausible.

Topics in Ethics – MT17 – Hilary Greaves and Teru Thomas

Week 8: Emotion and rationality in ethics

1. Background: The Kantian project, and discontents
    a. Morality is grounded in 'reason'/rationality: Necessarily, an agent who acts immorally is irrational.
    b. How this is supposed to work, in Kant's hands
        i. For any action, identify the 'maxim' that would be driving that action.
        ii. 'Contradiction in conception' test: Is there any possible world in which *everyone* acts on this maxim?
        iii. 'Contradiction in the will' test: If the maxim passes the first test: Is the possible world in question one that the agent can rationally 'will', given the motivation involved in the maxim itself?
        iv. If a given maxim fails either test, acting on that maxim is irrational.
        v. This is supposed to rule out, e.g.: lying, murder, suicide, never helping anyone else.
    c. Basic obstacle for the Kantian project (and the moral rationalist project more generally)
        i. Insofar as rationality is only about internal consistency, it cannot rule out a perfectly internally coherent 'upside-down morality'.
        ii. One response is to appeal to a more 'substantive' notion of rationality, so that the content of morality is built into rationality 'by definition'.
            1. But then what is the point of the rationalist claim? (E.g.: why be *rational* in that sense, rather than rational*?
2. Korsgaard on content scepticism vs motivational scepticism
    a. Content scepticism: Morality cannot be based on reason/rationality alone, because considerations of reason/rationality cannot impose any substantive constraints in the context of choice/action. (As above)
    b. Motivational scepticism: Morality cannot be based on reason/rationality alone, because
        i. Moral judgments are intrinsically motivating ('motivational internalism')
        ii. Reason/rationality alone cannot motivate (the 'Humean theory of motivation': motivation is generated by beliefs *and desires* working together, and reason/rationality guides only beliefs)
    c. Korsgaard's claim: Motivational scepticism presupposes content scepticism.
        i. Because: if considerations of reason/rationality *can* impose substantive constraints, that is to say that they can govern the choice of ends as well as the identification of means, so (a crucial component of) the Humean theory of motivation is false.
3. Some empirical results on the roles of the emotions in the practice of morality
    a. Moral dumbfounding (Haidt, 'The emotional dog and its rational tail')
        i. Moral judgments are caused by emotions. People try to provide post-hoc rationalisations of their judgments, but the judgments are not in fact caused by those rationalisations, nor are they typically revised when people realise the rationalisations don't make sense/don't apply to the case at hand.
    b. Hypnotism studies

        i. Subjects hypnotically conditioned to feel disgust at such words as ‘take’ and ‘often’ judge completely innocuous actions that are described using those words to be morally wrong. (Haidt and Bjorkland)
    c. Empathy and altruism (Batson)
        i. Empathy, roughly: experiencing an emotion of the same valence as the one a patient is perceived to feel (e.g. experiencing distress at witnessing others suffering)
        ii. People exhibit more genuinely altruistic behaviour when they feel empathy for the patient.
    d. fMRI studies (e.g. Greene)
        i. In the brains of subjects making *some* moral judgments, the most active areas of the brain are those involved in processing emotions, not those involved in reasoning.
            1. Greene: This is so for ‘deontological’ moral judgments, but not for utilitarian ones. (But see Kahane et al for dissent.)
4. Some types of moral sentimentalism
    a. General theme of sentimentalism: ‘Emotions are central to morality’
    b. 4 types of sentimentalism (Kauppinen)
        i. Explanatory sentimentalism: Emotions explain moral judgments
        ii. Judgment sentimentalism: Moral judgments are constituted by emotions (~non-cognitivism?)
        iii. Metaphysical sentimentalism: Moral facts are facts about what causes [[or merits]] emotional responses
        iv. Epistemic sentimentalism: Emotions are basic sources of moral knowledge/justification
    c. Basic question: When emotions influence moral judgment, to what extent should we regard these influences as *distortions*, vs. *sources of moral insight*?
        i. Explanatory sentimentalism is not the same as metaphysical/epistemic sentimentalism. (Not always appreciated in the experimental literature?)
5. Slote’s sentimentalist theory
    a. Slote’s basic claim: Normal emotional reactions of ‘second-order-empathy’ ‘fix the reference of’ such terms as “morally right” and “morally wrong”.
    b. Second-order empathy
        i. First-order empathy: experiencing e.g. distress at witnessing the suffering of another.
        ii. Second-order empathy: experiencing emotions of ‘warm approval’ (resp. ‘cold disapproval’) at witnessing others acting in ways that would be motivated (resp. would go against) first-order empathy.
    c. Reference-fixing: background
        i. Suppose I define ‘one metre’ by the stipulation: “One metre is the length of the standard bar held in Paris on January 1, 2017.”
        ii. Some questions (and answers):
            1. Can I know a priori that the standard bar is 1 metre long on 1.1.17?
                a. No. What I can know *a priori* is (only) that the *sentence* “One metre is the length of the standard bar held in Paris on January 1, 2017” is true. But I cannot know a priori which proposition is expressed by this sentence.

2. Could the standard bar have been other than one metre long on 1.1.17?
    a. Yes. What could not have happened is that the above *sentence* be false (holding fixed that the sentence is used as a stipulative definition). But the sentence could have had a different meaning. If the standard bar had had some other length, then "one metre" would have referred to a length other than one metre. (Compare: In French, "court" does not refer to courts.)

iii. What is going on
1. The above stipulation *fixes the reference* of the phrase 'one metre'.
2. This is distinct from the description "length of the standard bar held in Paris on January 1, 2017" *having the same meaning* as the phrase "one metre".
    a. If they had the same meaning, they would have the same reference in all possible worlds (not only in the actual world), so the answers to the above two questions would be different.
3. Reference-fixing
    a. There is some property that satisfies the given description in the actual world.
    b. But we have a privileged criterion of transworld identity for properties, and that privileged criterion does not coincide with: satisfaction of the given description. (In this example, the privileged criterion is: being the same length.)
    c. When we evaluate sentences containing the term in question (here, 'one metre') at other possible worlds, we take that term to refer to the *same property it refers to in the actual world*, not to *whatever satisfies the given description in the other possible world*.
    d. What determines the reference of the term (for the purpose of evaluation at any possible world) is: The description, together with the facts about which property *actually* satisfies that description, and the privileged criteria of transworld identity for properties.

d. Slote's reference-fixing proposal
    i. The proposal (recall): Normal emotional reactions of second-order-empathy fix the reference of such terms as "morally right" and "morally wrong".
    ii. Therefore: "morally right" (resp. "morally wrong") refers to actions of kind K iff actions of that kind *actually* elicit second-order empathic reactions of warm approval (resp. cold disapproval) in normal subjects.
    iii. Therefore: actions of kind K are morally right (resp. wrong) iff they actually elicit second-order empathic reactions of warm approval (resp. cold disapproval) in normal subjects.

e. A misguided objection
    i. "Slote's theory implies that if we had had warm-approval reactions to acts of murder, then murder would have been morally right. But that is absurd."

    ii. Reply: Slote's theory does not have this implication (just as: if the standard bar had been twice as long as it actually is, it wouldn't have been one metre long).
f. Classifying Slote's theory
    i. Most directly seems to be "semantic sentimentalism": emotions fix the reference of moral terms.
    ii. Relatedly, seems to imply epistemic sentimentalism; less clear relationship to the other 3 types above (?)
g. Objection: This theory is ultra-conservative of actual (current?) emotional reactions. Surely we shouldn't be that confident that none of our actual emotional-reaction tendencies is misguided?
h. Application: Prioritising identified over statistical lives