# Rationality & Ethics

Ryan Doody
Office: 6723
ryan.doody@mail.huji.ac.il

## 1 Course Description and Objectives

If you're instrumentally rational, you should take the best means to your ends. But what should you do when you don't know which means is best? If you're moral, you should do what is morally best. But what should you do when you don't know what morality requires of you? This course will explore some possible answers to these questions. The first part of the course will introduce the tools of bayesian decision theory, examine its foundations, and discuss some puzzles. The second part of the course will evaluate the prospects of applying these tools to ethics. What should we do when we are unsure about the consequences of our decisions? What should we do when we are uncertain about which moral theory is correct? Is expected moral value always well-defined?

The course is broken up into four sections:

(1) Introduction to Decision Theory

(2) Well-Being & Time

(3) Aggregating Well-Being across different people

(4) Moral Uncertainty

The first part (Lectures 1-6) will introduce the formal tools of decision theory, explore its axiomatic foundations, and address some related puzzles (e.g., the Newcomb Problem, the Allais Paradox). The second part (Lectures 7-9) is concerned with the well-being of a person at a time, the well-being of a person over time, and the relationship between the two. The third part (Lectures 10-12) is about aggregating the well-being of individual people to arrive at the overall value of a state-of-affairs. We will evaluate arguments for Utilitarianism (e.g., Harsanyi, Broome), investigate the role that uncertainty plays in these arguments, and look at some puzzles concerning Population Ethics. The final section (Lectures 13-14) will be about moral uncertainty.

## 2 Requirements

You should come to class each meeting having carefully read the material and ready to participate. The best way to learn about philosophy is to *do* philosophy. We will be engaged in the project of all doing philosophy together.

## 3 Assignments

You are required to write a term paper for the course. Throughout the term, I encourage you to meet with me (at least once) about the paper.

## 4 (Tentative) Schedule

### Part I: Practical Rationality

- **Meeting 1: Introduction to Decision Theory & Expected Value**
  - No Readings

  *Decision Theory is about how people make — and should make — decisions (especially under conditions of risk or uncertainty). Of particular interest is the view — called Expected Utility Theory — which says that rational agents should take the option, out of those available, that maximizes expected utility. We'll get clearer about what this means.*

- **Meeting 2: Subjective Expected Utility Theory, Savage vs Jeffrey**
  - Chapter 2, Chapter 3.1–3, and Chapter 5 of Leonard J. Savage, *The Foundations of Statistics.* Dover Publications. New York, NY. 1954
  - Chapter 1, Chapter 2, and Chapter 5 of Richard Jeffery, *The Logic of Decision.*

  *What, exactly, is 'expected utility' and why should we maximize it? We'll look at two competing views of Subjective Expected Utility Theory. Both views set out a number of constraints on rational preference and show that, if your preferences obey these constraints, you can be represented as maximizing expected utility. (These are called Representation Theorems.) We'll investigate the importance of these results, and discuss the ways in which these two views differ.*

- **Meeting 3: The Newcomb Problem**
  - Robert Nozick, "Newcomb's Problem and Two Principles of Choice," in *Essays in Honor of Carl G. Hempel,* Rescher ed., 1969
  - David Lewis, "Why Ain'cha Rich?"
  - Andy Egan, "Some Counterexamples to Causal Decision Theory," *The Philosophical Review.* 2006

  *Jeffrey's view seemed like an improvement over Savage's, but it appears to get the wrong result in the Newcomb Problem. Does it? We'll discuss the difference between Evidential and Causal Decision Theory. Which view is right?*

- **Meeting 4: Transitivity**
  - Larry Temkin, "A Continuum Argument for Intransitivity," *Philosophy and Public Affairs,* 25. 1996

*Your preferences can be represented with a utility-function only if they are* transitive *(if you prefer $X$ to $Y$, and you prefer $Y$ to $Z$, then you prefer $X$ to $Z$). Is it a rational requirement that your preferences be transitive? Temkin argues that the betterness-relation is intransitive. If so, it appears that it can be rational to have intransitive preferences.*

- **Meeting 5: Incommensurability & Opaque Sweetening**
  - Ruth Chang, "Hard Choices," *Journal of the Philosophical Association,* 3(1). 2017
  - Caspar Hare, "Take the Sugar," *Analysis*. 2010
  - Miriam Schoenfield, "Decision Making in the Face of Parity," *Philosophical Perspectives.* 2014

  Optional Readings:

  - John Broome, "Is Incommensurability Vagueness?" In *Ethics out of Economics,* Cambridge University Press. 1999
  - Ryan Doody, "Parity, Prospects, and Predominance," manuscript. 2017

  *Your preferences can be represented with a utility-function only if they are* complete *(for all items, $X$ and $Y$, you either prefer $X$ to $Y$, you prefer $Y$ to $X$, or you are indifferent between the two). Is it a rational requirement that your preferences be complete? Almost no one thinks so. It can be rational to have incomplete preferences if, for example, values are incommensurable. Is there a way of generalizing Expected Utility Theory to handle cases of incomplete preferences? This gives rise to the problem of Opaque Sweetening.*

- **Meeting 6: The Allais Paradox & Risk-Aversion**
  - Chapter 1, Chapter 4, and Chapter 5.4 of Lara Buchak, *Risk and Rationality,* Oxford University Press. 2013

  Optional Readings:

  - Chapter 5.2–5.8 (and Chapter 4) of John Broome, *Weighing Goods: Equality, Uncertainty and Time,* Blackwell Publishers. 1991
  - Chapter 2 of of Lara Buchak, *Risk and Rationality,* Oxford University Press. 2013

  *Risk-aversion — as illustrated in the Allais Paradox — is incompatible with Expected Utility Theory. Is it irrational to be genuinely risk-averse? Buchak argues that it isn't, and offers an alternative view.*

## Part II: Well-Being & Time

- **Meeting 7: Well-Being**
  - Shelly Kagan, "The Limits of Well-Being," *Social Philosophy & Policy,* vol. 9 (2). 1992
  - Chris Heathwood, "The Problem of Defective Desires," *Australasian Journal of Philosophy,* vol. 83 (4). 2005

*We'll look at various theories of well-being: of how things may be better or worse for a person at a time. Kagan provides a helpful taxonomy, and evaluation, of the canonical proposals. Heathwood offers a defense of the* actual desire satisfaction theory *(which says that things go well for you to the extent that your actual desires are satisfied).*

- **Meeting 8: The Shape of a Life**
  - David Velleman, "Well-Being and Time," *Pacific Philosophical Quarterly*. 1991

  *How does your well-being at the particular moments in your life relate to the value of your life overall? Does the* distribution *of the good and bad moments throughout your life (your momentary well-being) affect your life-time well-being? Velleman argues that it does. In fact, he argues that the facts about your life-time well-being don't supervene merely on facts about your momentary well-being. We'll discuss whether or not he's right.*

- **Meeting 9: Time-Bias**
  - Derek Parfit, "Different Attitudes to Time," Chapter 8 of *Reasons and Persons,* Oxford University Press. 1984
  - Caspar Hare, "Time — The Emotional Asymmetry," in *A Companion to the Philosophy of Time,* ed. Adrian Bardon and Heather Dyke, Wiley Blackwell. 2013

  Optional Reading:
  - Caspar Hare, "A Puzzle About Other-Directed Time-Bias," *Australasian Journal of Philosophy*, Vol. 86 (2). 2008

  *Does it make sense to care about whether good moments are in our pasts or our futures? Is it irrational to prefer that pain be in our past and pleasure in our future? Is it irrational to discount the far-off future more heavily than the very-near future? Perhaps.*

## Part III: Aggregation Across People

- **Meeting 10: Weighing the Interests of Infinitely Many**
  - Shelly Kagan and Peter Vallentyne, "Infinite Value and Finitely Additive Value Theory," *Journal of Philosophy,* Vol. 94 (1). 1997
  - Frank Arntzenius, "Utilitarianism, Decision Theory and Eternity," *Philosophical Perspectives,* 28, *Ethics*. 2014

  *The world might contain an infinite number of things of moral concern. How should we weigh these (potentially) infinitely many interests off of each other? If the overall value of the universe is either infinitely good, infinitely bad, or undefined — and if your actions are only capable of adding or subtracting, at most, a finite amount of value to the total, is everything morally permissible?*

- **Meeting 11: Ethics & Uncertainty 'Behind the Veil'**
  - John Harsanyi, "Morality and the Theory of Rational Behavior," *Social Research* 44. 1977

- Chapter 3.3 of John Broome, *Weighing Goods: Equality, Uncertainty and Time,* Blackwell. 1991
- Caspar Hare, "Should We Wish Well to All?" *Philosophical Review*, 125 (4). 2016

Optional Reading:

- Lara Buchak, "Taking Risks Behind the Veil of Ignorance," manuscript
- John Broome, "Utilitarianism and Expected Utility," *Journal of Philosophy,* 84 (8). 1987
- Marc Fleurbaey and Alex Voorhoeve, "Decide as You Would with Full Information! An Argument against ex ante Pareto," In: Eyal, Nir and Hurst, Samia A. and Norheim, Ole F. and Wikler, Dan, (eds.) *Inequalities in Health: Concepts, Measures and Ethics. Population-level bioethics,* Oxford University Press. 2013

*Some moral theories — those, for example, that give weight to fairness in the distribution of well-being — entail that we should not always do what is expectedly best for all. Is this a problem for those theories? Harsanyi and Hare argue that it is. We'll discuss whether they are right.*

- **Meeting 12: Weighing Lives, present and future**

  - Chapter 16–19 of Derek Parfit, *Reasons and Persons,* Oxford University Press. 1984
  - Excerpts from John Broome, *Weighing Lives,* Oxford University Press. 2004

  *Our actions not only affect the well-being of those alive now, they also affect the well-being and identities of those who will exist in the future. How should we weigh the interests of those alive now against the interests of those who might exist in the future?*

## Part IV: Normative Uncertainty

- **Meeting 13 & 14: What to do when you don't know what you should do**

  - Johan E. Gustafson & Olle Torpman, "In Defense of My Favourite Theory," *Pacific Philosophical Quarterly,* 95 (2). 2014
  - Brian Weatherson, "Running Risks Morally," *Philosophical Studies,* 167. 2014
  - Ittay Nissan-Rosen, "Against Moral Hedging," *Economics and Philosophy,* 31. 2015
  - Andrew Sepielli, "What to do when you don't know what to do when you don't know what to do..." *Nous*, 48 (3). 2014

  *We've looked at questions concerning what you should do (both rationally and morally) under conditions of non-normative uncertainty — that is, in cases in which you are unsure of how the world actually is, moral (and other normative) facts aside. But what about cases of* normative *uncertainty? What should you do when you don't know what's required you? Should we treat normative uncertainty the same way we treat non-normative uncertainty?*