

Philosophy 416
Winter 2012
Uses and Abuses of Game Theory in Philosophy

Instructor: David DeVidi

Time: Tues & Thurs 11:30—12:50

Place: AL 210

Office Hours: Tues 10—11:15, or by appointment

“Decision theory ... represents one of the great scientific achievements of all time, beginning with Bernoulli and Pascal in the seventeenth and eighteenth centuries, and culminating in the work of Ramsey, de Finetti, Savage and Von Neumann and Morgenstern in the early and middle years of the twentieth century.” So says Herbert Gintis. He is one of the pre-eminent authors currently working in decision and game theory, so perhaps he is biased. But decision theory is undoubtedly profoundly important because of its wide range of applications, not only for the purposes for which it was first created but as a foundation of other areas of research.

Traditional decision theory was primarily concerned with the question of what constituted a rational choice for an individual agent with particular preferences and particular beliefs about the circumstances in which an action would take place. Later, especially during World War II, it became clear that it was important to have tools available that took explicit account of an important feature of the environments in which decisions are made---the presence of other decision makers. Game theory is a formal theory of rational decision making where more than one rational agent is involved.

Since game theory in its earliest form was developed as a tool for analyzing strategic situations in a time of war, it was by some regarded as of limited use because it was thought to involve idealizations that rarely had much to do with actual decision makers: for instance, idealized agents who are assumed to be fundamentally selfish and to have enormous (perhaps infinite) capacities to store information and to crunch numbers. Recent developments have shown this caricature to be quite mistaken, and game theory in various forms is now fundamental to a variety of fields, including evolutionary biology, economics, and some parts of political theory.

Both traditional decision theory and game theory have also played an important role in various parts of philosophy, including in moral theory, logic, epistemology and the philosophy of science: so much so that most good philosophy departments offer courses in decision theory and game theory (and so these subjects have themselves become part of philosophy). The role they have played in philosophy, alas, has not always been for the good.

In this class we will look at a few discussions in the philosophical literature that involve decision theory and, especially, game theory. However, to understand these discussions, it will be necessary to have a reasonable grasp of the bits of decision theory and game theory involved. The course will therefore

Philosophy 416

involve learning of two sorts: (1) getting a clear grasp on the formal tools of decision theory and game theory; and (2) sorting out, and eventually being able to participate in, some of the philosophical discussions that involve these tools.

The basic format for the course will be to switch back and forth between learning something about the formal apparatus of decision theory and game theory and more discursive philosophical discussions. In a rough and ready way, the goal will be to cover enough of the formal material to allow students to understand the philosophical discussions to follow.

The class will be made up of students with a variety of backgrounds. Most, but not all, will be philosophers. It is therefore not possible to have a reasonable “textbook” to cover the formal material we need. Some of the books out there are all either too simple or too narrow to allow us to deal with all the topics we’ll be running up against. Most of the others are aimed at senior students with more mathematical background than it’s reasonable to expect from philosophy undergraduates. Fortunately, often the extra complexity is a result of presenting the material “in full generality,” so that what are often ideas that are fairly easy to grasp can get lost in a blizzard of symbols if you’re not used to reading mathematics. For this course, the formal material will therefore be presented in course notes and discussed in the lectures using only the degree of mathematical sophistication necessary to grasp the key ideas.

You will have trouble in this class if you are very math-phobic, but the mathematical prerequisites will not go much beyond high school algebra. If you have already done well in Philosophy 216 or in a first course in logic, you should be fine. However, since we will not be using a textbook, **it is very important to attend the classes: if you don’t do so you may find yourself quickly falling behind.**

As you’ll see below, the evaluation for the course will be based both on your grasp of the formal tools and on your ability to write sensibly about it. People with a bit of formal background will obviously have an advantage in one part; people with a lot experience writing philosophical essays will obviously have an advantage on the other. My hope is that all students will be able to improve in both parts, but especially in the part that they don’t already have a lot of background in.

Besides lecture notes and other materials I’ll be putting on the LEARN site for the course, the course readings will all be journal articles available on line, and accessible with minimal difficulty using the UW Library web site. This approach makes sense because while there are many collections of papers on decision theory and game theory, they are all much too tightly focused for our purposes---that is, they tend to focus on one debate, while we’re going to touch on a few different debates. I’m guessing that you won’t have much trouble accessing the articles, but if you do you can get in touch with me or with the liaison librarian for philosophy at the UW Library, Christine Jewell, at cjewell@uwaterloo.ca.

This is the first time this course has been offered. This means that it’s a bit difficult to predict how quickly we can get through material. And anyway, for classes that attract a varied group of students, it’s important to set the pace so that all the students who are putting in a real effort are learning the stuff---I’d rather teach the students actually in the class than follow some pre-set idea of what the students ought to be like. That’s why there are no specific dates attached to each section. The material I have

Philosophy 416

mapped out might take us through the whole term, or we might have a couple of weeks left at the end. If we have more time, we will choose another topic based on student interest from among many possibilities, for instance multi-agent modal logic or social epistemology.

Evaluation:

There will be **three in-class tests**, each worth 20%, and a term paper, of 6-8 pages, worth 40%. The tests will take place on **Jan 19, Feb 16 and March 22**. The **term paper** is due on the last day of class, **Mar 29**.

Obviously, none of the evaluation is based on group work. But in addition to the course notes, I will be providing occasional problem sets or exercises (e.g., there are some in the first set of notes) that you can work on to make sure you're grasping the technical material. These provide a chance for you to work together with your classmates, if that's a productive way for you to learn. We'll take up some of these problems in class. But these questions are relevant to your evaluation only because doing them will make you much better prepared for the tests.

Attendance and participation. Regular attendance, and being prepared to take part in discussions, is essential to getting full value out of this course. Indeed, the regular attendance and preparation to participate of the other members of the class is also essential for you to get as much out of the class as possible. There will not be a "class participation" mark, but we will be taking up problems from the problem sets and discussing the philosophical material, and coming prepared for this will greatly increase your chances of doing well on the tests. I'm counting on the fact that you've been in university a while to ensure that you're able to make an adult decision about attending class rather than needing to use participation marks to incentivize it. By the end of the course, you may have tools to evaluate whether this is a stupid move on my part or not!

Schedule:

Part 1: Decision Theory (and its discontents?)

Readings: Course notes, to be distributed (via Waterloo LEARN).

Loomis, Graham, "When Actions Speak Louder than Prospects", *American Economic Review* 78, 3 (1988) 463-70.

Kahneman, Daniel, Jack L. Knetsch and Richard H. Thaler, "Anomalies: The Endowment Effect, Loss Aversion and Status Quo Bias," *Journal of Economic Perspectives* 5, 1 (1991) 193—206.

Part 2: Basics of Game Theory

Readings: Course notes.

Part 3: Game theory, cooperation and altruism

Readings: Course notes.

Philosophy 416

Sober, Elliot, "The Evolution of Altruism: Correlation, Cost and Benefit," *Biology and Philosophy* 7 (1992) 177-87.

Carter, Alan, "Evolution and the Problem of Altruism," *Philosophical Studies* 123 (2005) 213-30 .

Part 4: Game theory and justice

Readings: Course notes.

Skyrms, Brian, "Sex and Justice," *Journal of Philosophy* 91, 6 (1994) 305-20.

Vanderschraaf, "Game Theory, Evolution and Justice," *Philosophy and Public Affairs* 28, 4 (2000) 325-58.

Part 5: Game theory and rationality: equilibrium refinements, common knowledge, etc.

Readings: Course notes.

Gintis, Herbert, et al., "Strong Reciprocity and the Evolution of Human Morality," *Social Justice Research*, 21 (2008) 241—53

Gintis, Herbert, "Social Norms as Choreography," *Politics, Philosophy and Economics*, 9, 3 (2010) 251—264.

Collaborative work, acknowledging sources and avoiding plagiarism: While we shall endeavor to maintain a collaborative atmosphere in our in-class discussions, and you may find it useful to work together with classmates on parts of the problem sets, or in reviewing the class notes, or to discuss the readings. But all the material submitted for evaluation in this class is submitted with the understanding that it is your own, individual work. I hope this goes without saying for the tests! For the term paper, this means that when you begin to write your paper your discussion of the particular topic of the paper with classmates and others stops until after you've handed the paper in. I encourage you to finish a draft well before the deadline, so you can set it aside for a day or two then go back to proofread it. Do not have your classmates, parents, friends, roommates, or anyone else do the proofreading for you.

Referencing of sources used during your research for your papers should accord with professional philosophical standards. This does not mean you need to list every paper you've ever read, nor that you need to cite things that are plausibly "common knowledge." But direct quotations need to be sourced, and if you've borrowed an idea that a reader might be misled into thinking is one you came up with on your own, you need to say where it's from. We'll discuss this more in class, and if you are in doubt, ask the course instructor for advice about whether something needs to be cited or not.

Any standard format for references is acceptable, though some version in which citations are "in text" and footnotes are reserved for more substantial comments, such as the APA style, is my preference. A brief introduction to the APA style, if you don't know it, can be found in either of these places:

<http://library.concordia.ca/help/howto/apa.php>

Philosophy 416

<http://ereference.uwaterloo.ca/display.cfm?categoryID=15&catHeading=Citation%20/%20Style%20Guides#AmericanPsychologicalAssociation%28APA%29>

Late work: As a general policy, late work will be penalized. The minimum penalty is a 10% deduction from the mark earned on the assignment, though work more than one week late will be subject to a 20% deduction, and work more than two weeks late will receive a mark of zero. These penalties will apply to any un-excused late work. Excuses must be arranged with the instructor as soon as it becomes clear to the student that a piece of work will be late and the excuse is legitimate. In all cases, including serious illness, appropriate documentation establishing the truth of the excuse is required---the reason for this rule is simply to remove the need for me to make judgements about whether I believe a particular student or not, and so risk unfairly demanding proof from some students but not others. For illness, normally a doctor's note or similar documentation is required (and not, as one student once produced for me, a letter from your mom).

The Faculty of Arts requires that the information below appear on all course outlines. Please be sure you are familiar with it.

Cross-listed course:

Please note that a cross-listed course will count in all respective averages no matter under which rubric it has been taken. For example, a PHIL/PSCI cross-list will count in a Philosophy major average, even if the course was taken under the Political Science rubric. Page 2 of 7

Academic Integrity:

Academic Integrity: In order to maintain a culture of academic integrity, members of the University of Waterloo are expected to promote honesty, trust, fairness, respect and responsibility.

Discipline: A student is expected to know what constitutes academic integrity, to avoid committing academic offences, and to take responsibility for his/her actions. A student who is unsure whether an action constitutes an offence, or who needs help in learning how to avoid offences (e.g., plagiarism, cheating) or about "rules" for group work/collaboration should seek guidance from the course professor, academic advisor, or the Undergraduate Associate Dean. When misconduct has been found to have occurred, disciplinary penalties will be imposed under Policy 71 – Student Discipline. For information on categories of offenses and types of penalties, students should refer to Policy 71 - Student Discipline, <http://www.adm.uwaterloo.ca/infosec/Policies/policy71.htm>

Grievance: A student who believes that a decision affecting some aspect of his/her university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70 - Student Petitions and Grievances, Section 4, <http://www.adm.uwaterloo.ca/infosec/Policies/policy70.htm>

Appeals: A student may appeal the finding and/or penalty in a decision made under Policy 70 - Student Petitions and Grievances (other than regarding a petition) or Policy 71 - Student Discipline if a ground for an appeal can be established. Read Policy 72 - Student Appeals, <http://www.adm.uwaterloo.ca/infosec/Policies/policy72.htm>

Philosophy 416

Academic Integrity website (Arts): http://arts.uwaterloo.ca/arts/ugrad/academic_responsibility.html

Academic Integrity Office (UW): <http://uwaterloo.ca/academicintegrity/>

Accommodation for Students with Disabilities:

Note for students with disabilities: The Office for Persons with Disabilities (OPD), located in Needles Hall, Room 1132, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with the OPD at the beginning of each academic term.