

Teaching Dossier

Milan Mossé

Contents

Teaching Statement	2
Diversity Statement	3
Teaching Experience	4
Summary of Qualitative Student Evaluations	5
Summary of Quantitative Student Evaluations	10
Course Descriptions	11
Syllabus Preliminaries	14
Syllabus: Bioethics	17
Syllabus: Ethics and Artificial Intelligence	19
Syllabus: Introduction to Moral and Political Philosophy	22
Syllabus: Philosophy of Science	25

Teaching Statement

I bring a bag of tangerines to class and throw one to any student who asks a question and wants one. To help students review before a midterm, I bring buzzers and run a Jeopardy-style game, where teams of students choose the topic and difficulty of each question and earn points for answering it correctly. I dress up on Halloween, and I encourage my students to do the same. We watch clips from movies that illustrate concepts from the readings. As students walk in, we talk with each other about how the week is going, and about how they are finding the class.

I want my students to look forward to class every week, and I want them to feel comfortable and relaxed. This is valuable to me in itself—students are stressed and over-worked, and they deserve to feel cared about as people. But it is also important to me that my students feel excited about coming to class and comfortable speaking up, because I know it will allow them to do their best work.

In many ways, I am a serious teacher. I give pop quizzes on the readings. I cold call on small discussion groups after they have had time to talk among themselves. I require that students bring copies of the readings. And I take attendance in every class. These are somewhat strict policies, and without the proper context, they can intimidate students. These policies are effective because I build trust with my students, partly by introducing games and snacks that make class feel light-hearted, and partly by softening my policies in ways that do not detract much from their efficacy. For example, students pair up to compare answers on quizzes before turning them in, and they are graded with a threshold, rather than a letter grade. I always remind students that “no thanks—I’ll pass” is a totally OK answer in response to a cold call. I show students where they can print readings for free on campus (just in case, I always have extra copies of the readings for them to borrow). And I take attendance by greeting students by name as they come in—they know I’m taking attendance, but they also know I can only do it because I memorized their names, which I think helps them to feel seen as individuals.

Fair and holistic evaluation is essential to my teaching. I reward hard work: students get extra credit points for showing up in office hours, and for doing optional quizzes and readings. I give my students opportunities to revise their written work after receiving feedback. I design difficult assignments that challenge them, but I grade with respect for my students’ effort and insight. This requires reading several submissions before I begin grading, to learn about the correct solutions that my students discovered and that I hadn’t anticipated. It requires working to see how a student is on the right track, so that they are recognized with credit for what they get right. And it requires curving grades up when I simply made the assignment too hard. When grading is transparent and emphasizes students’ successes, it helps them to internalize high standards of clarity and interpretive charity, and gives them the confidence to defend their own views.

A central goal of mine is to help students to develop their philosophical voices. They write frequently and practice rephrasing ideas in their own words. When we discuss [my writing guide](#), we talk about how style and argument interact. Finally, my students shape the course: I collect anonymous feedback, show them the results, and explain exactly what I’m changing in response. We learn together.

Diversity Statement

Diversity, equity, and inclusion structure my approach to teaching, mentorship, departmental citizenship, and research. Here are three things I do in the classroom to foster an inclusive learning environment. First, as you'll see in the attached syllabi, I am a strong proponent of syllabus diversification. Second, especially out of respect for neurodiversity, I communicate lesson plans before the start of class, use visual aids to illustrate key concepts, and ensure that my students have a variety of ways to participate—online discussion posts, office hours, quizzes, worksheets, and small group work all count toward the participation grade, and we discuss [different ways to contribute](#) to a philosophical conversation. Third, I co-authored Berkeley's [current philosophy department guidelines](#) for trans and nonbinary inclusion in philosophy, and I follow these in my own classroom.

Mentoring students is a core priority for me, especially in supporting those from underrepresented backgrounds to thrive academically and professionally. At Berkeley, I have mentored undergraduates applying to philosophy PhD programs, and I have written several letters of recommendation in support of students' applications for scholarships and law schools. With Hannah DeBrine, I created, organized, and participated in the philosophy department's Directed Reading Program, which (running for three years now) pairs undergraduates with graduate student mentors to discuss weekly readings, culminating in undergraduate presentations. We emphasize outreach to under-represented groups, and almost all undergraduate participants come from gender, racial, or ethnic groups that are under-represented in philosophy.

I have also worked on several projects aimed at advancing diversity, equity, and inclusion in philosophy. For two years, I co-organized Berkeley's [Minorities and Philosophy](#) chapter. This involved organizing several events for undergraduates, including bi-weekly meetings with meals, reading groups on masculinity and Latin American philosophy, and Q&A sessions with undergraduate audiences and panels of faculty and graduate students. Madeleine Levac, Tim Crockett, and I drafted a [successful proposal](#) to un-name (Bernard) Moses Hall—which is now called Philosophy Hall. Virginia Foggo and I ran the department's first ever [COMPASS workshop](#), which helps undergraduates from groups that are under-represented in philosophy to apply to philosophy graduate programs.

Finally, my research directly addresses issues related to diversity, equity, and inclusion. With Frederick Eberhardt, Thomas Icard, and Kara Schechtman, I continue to work on [papers](#) that provide causal models of discrimination. In a paper which builds on my dissertation, **Against Deportation**, I argue that host countries have duties of reciprocity not to deport immigrants who reside and work illegally: states cannot permissibly accept the benefits of immigrants' compliance with the most costly duties that apply to citizens, while withholding basic benefits like residence associated with citizenship. This topic carries personal significance for me, as a second-generation Argentine immigrant to the United States.

Teaching Experience

University of California, Berkeley

Fall 2025	TA, Philosophy of Science, with Ezra Rubenstein
Fall 2024	TA, Political Philosophy: Equality and Identity, with Véronique Munoz-Dardé
Sum 2024	Co-Instructor , Individual Morality and Social Justice, with Russell McIntosh
Fall 2023	TA, Political Philosophy, with Véronique Munoz-Dardé
Sum 2023	TA, Contemporary Ethical Issues, with Tim Crockett
Fall 2022	TA, Bioethics, with Johann Frick
Sum 2022	TA, Philosophy of Race, Ethnicity, and Citizenship, with Tim Crockett
Spr 2022	TA, Individual Morality and Social Justice, with Johann Frick
Fall 2021	TA, Nature of Mind, with John Campbell

Stanford University

Spr 2024	TA, Incentives in Computer Science, with Aviad Rubenstein
Sum 2020	TA, Design and Analysis of Algorithms, with Karen Shi
Win 2020	TA, Design and Analysis of Algorithms, with Mary Wootters
Win 2018	TA, Programming Abstractions, with Marty Stepp
Fall 2018	TA, Programming Methodology, with Mehran Sahami

Summary of Qualitative Student Evaluations

Teaching is the most rewarding part of my academic life. I'm grateful that this is reflected in my students' evaluations:

Something distinctive about Milan is that he offers a lot of support for his section! He provided us with study guides, extra office hours and was very responsive with emails! He's the **most attentive GSI I've ever had** and I appreciate all the effort he put in. He also made the section inclusive for everyone. There was nothing unhelpful. (Nature of Mind, Fall 2021)

There is honestly nothing I would change. **This has been the best section I've been a part of** since we not only get to interact with our peers but also always leave with a clear understanding of everything. (Individual Morality and Social Justice, Spring 2022)

Milan was really helpful in reviewing weekly materials and essay preparation. I could really tell he cares a lot and takes section prep seriously. OH was really helpful and practical too. He was really helpful with 1:1 questions, and would be instructive in listening to me and pointing out which parts I'm confused on. Milan was **one of the best GSIs I've had at Berkeley**, would love to be in another course assisted by him! (Contemporary Ethical Issues, Summer 2023)

Extensive office hours were helpful. I most enjoyed how Milan moved between small-group and class discussion and provided ample time for students to try to understand examples and discuss how they evaluate arguments. Overall one of the **best GSIs I have had**. (Political Philosophy, Fall 2023)

It's a really well-organized, logical and systematic course, probably the best course I've ever taken. Also topics are appealing. Readings are challenging but fun. (As Primary Instructor for Individual Morality and Social Justice, Summer 2024)

Milan is an amazing GSI/instructor and has been my favorite one since being at UC Berkeley. I'm confident he will continue to do so. (As Primary Instructor for Individual Morality and Social Justice, Summer 2024)

After TA-ing Bioethics in Fall 2022, I won the Outstanding GSI (Graduate Student Instructor) Award, granted to fewer than 10% of TAs at Berkeley. Student from that semester write:

Favorite GSI so far.

Absolutely my favorite GSI in all my classes here!

Milan was FANTASTIC. If I could have him as my GSI again, I would without hesitation.

To be clear, this is the **first time I've given all 7's** on a GSI evaluation. [7 = highest possible score.]

Milan is one of the best GSIs at Berkeley. Some particularly distinctive things are: 1) The way he treats students' participation/questions/answers: reconstructing them in a charitable way and being respectful even when conversation becomes a bit charged, as they sometimes do with controversial topics. 2) Availability outside of class. Milan has review sessions for challenging material, and even 1-on-1 appointments for papers. 3) Written feedback was substantial and extensive, detailed and helpful.

The last student mentions three features of my teaching—I'll take each in turn.

1) Participation:

I really like how Milan goes out of his way to not only be understanding and accommodating, but also how it **never felt like he was just depositing knowledge onto us. We felt like philosophers and formed our own thoughts** and applications of the material. (Nature of Mind, Fall 2021)

The sections were well-structured. Also, despite having an agenda, **Milan would focus on areas the group had concerns about, allowing us to explore those ideas** instead of sticking outright to the agenda. Much appreciated. (Philosophy of Race, Ethnicity, and Citizenship, Summer 2022)

Milan is really organized with sections. **He always asks students which topics we'd like to focus on and lead discussions accordingly.** I especially found the section on philosophical writing to be particularly helpful. (Philosophy of Race, Ethnicity, and Citizenship, Summer 2022)

One thing I really liked about Milan's discussions was that **whenever a student would raise a point/idea or comment, he would lay it out and rephrase it again for the rest of the class so that everyone could be on the same page about it.** This I believe encouraged more students to participate in the discussion and resulted in more fruitful conversations. (Bioethics, Fall 2022)

I really enjoyed the tone set by Milan in section. He crafted great lesson plans, budgeted our hour timeline really well, and made the learning of dry material way less dry. I found his presentations simple, effective, and natural. (Bioethics, Fall 2022)

Milan was great at **charitably translating questions** into very interesting extensions of what we were talking about, and was great at **adapting to what the section wanted to discuss.** (Bioethics, Fall 2022)

I always feel very engaged when Milan is teaching. His examples of hard-to-grasp concepts always made them digestible. **I loved the amount of time given for student discussion. I also appreciated how no question was a bad one.** (Bioethics, Fall 2022)

Milan encouraged participation in a fantastic way. He was very available to talk about material in and out of class, and he always ensured students understand the class concepts. I always find Milan extremely helpful. He's a great teacher and he will have a great future as a lecturer/professor if that's what he pursues. (Bioethics, Fall 2022)

The handouts were very helpful and clearly written. Review sessions during OH were also useful in writing essays. **I think Milan did an excellent job of encouraging participation, and making students feel comfortable.** (Contemporary Ethical Issues, Summer 2023)

2) Support:

Very accessible over email and in office hours, provided additional course material such as reviews, very quick turn around time with assignments and feedback, Milan could gauge our understanding of the material well and could tell when we needed more support. Overall, Milan was an amazing GSI and he was one of the best parts of this course. I looked forward to coming to section each week. (Nature of Mind, Fall 2021)

The sections were well-structured. Also, despite having an agenda, **Milan would focus on areas the group had concerns about, allowing us to explore those ideas** instead of sticking

outright to the agenda. Much appreciated. (Philosophy of Race, Ethnicity, and Citizenship, Summer 2022)

The most distinctive thing he did was **creating quizzes for us every week**, which other GSIs didn't have. This was extremely helpful in helping me understand the readings and lectures. (Individual Morality and Social Justice, Spring 2022)

I found my GSI **incredibly approachable, patient, and a great ally** when I struggled and just needed a little more help—especially in email exchanges. Great response time there and with really helpful feedback. (Contemporary Ethical Issues, Summer 2023)

He used simple terms to describe complex ideas and **cared/considered the feedback and input we had. Best GSI I've ever had!** (Political Philosophy, Fall 2023)

Milan held extensive office hours and was easy to reach outside of class which assisted in my development as a philosophy student overall, because he was able to address specific issues with my writing. Milan also **wrote handouts for every single reading we did in class**. These were incredibly detailed and clear. The additional resources he provided to students to ensure our success was above and beyond what I have experienced with any other philosophy GSI. (Topics in Political Philosophy: Equality and Identity, Fall 2024)

3) Writing:

Milan was always open for our feedback in ways he could improve the sections and would take our feedbacks very seriously. **I thought the lesson on the premise-conclusion format/list and sound vs. valid arguments was very helpful!** (Nature of Mind, Fall 2021)

The sections were incredibly helpful, especially for writing the papers. Also the review sessions felt essential to my understanding of the concepts that were discussed. The way the paper outlines were explained was very helpful. **Feedback on papers was very helpful as well**, and explanations were always clear. (Philosophy of Race, Ethnicity, and Citizenship, Summer 2022)

Milan always came in with an absurdly detailed plan for section which I appreciated. It made it so argument dissection/reconstruction was very fruitful. It was always very clear/logical how one thing flowed to the next. What was most distinct was just how detail-oriented and prepared Milan was. I haven't seen any professor or GSI in the department show this level of either of these qualities—I found it extremely helpful in giving me a better grasp for the arguments & dialectics. This was showcased not only in section through his argument reconstruction, but also the notes he provided for the various arguments. Lastly, **his handout on how to write a philosophy paper should be shown in all philosophy classes—it's that good!** (Contemporary Ethical Issues, Summer 2023)

Section was very helpful in getting clear on paper topics and understanding the basic structure of arguments. They also provided good opportunities to talk and figure out our own disagreements with the arguments. **Great comments on papers and ability to talk through arguments with students.** Can tell what misconceptions students have even when they can't articulate it themselves. (Contemporary Ethical Issues, Summer 2023)

Milan is an incredible teacher. He went above and beyond in helping students prepare for exams and understand course material. He provided detailed handouts breaking down the arguments of each paper. The handouts also offered fruitful questions for discussion. **He also made himself available for one-on-one appointments to discuss our final papers and**

offered detailed and invaluable feedback on our initial thoughts. (Topics in Political Philosophy: Equality and Identity, Fall 2024)

My students frequently use the phrase “above and beyond” in course evaluations. Here are some examples of this, in addition to the listed mentioned above:

Milan always went **above and beyond** to help his students and was always very understanding. One instance, I showed up to his office hours believing I had arrived at my scheduled time, but it turned out that I had scheduled [our meeting] for a week earlier and missed it. He was very understanding and took additional time out of his day to sit down with me and go over my essay even though his office hours had officially finished. He also created a tremendously helpful and comprehensive study guide which cleared up a lot of the confusing topics in the semester. The time and thought put into making it was very apparent and appreciated sincerely. (Nature of Mind, Fall 2021)

I really appreciated Milan’s approachability and the support he gave to all students. Milan went **above and beyond** in class and office hours to make sure everyone understood the material and got the help that they needed. Milan put a lot of time and energy into his students and it showed. (Individual Morality and Social Justice, Spring 2022)

Excellent examples, good at reading the room to gauge our tracking/understanding, seems to have a well thought-out lesson plan for every class to a degree much higher than my other GSIs, probably the best GSI I’ve had so far, wants students to understand the material and goes out of his way to help, spends way more time than any other GSI I’ve had on his “GSI duties” like emailing us back, section prep, office hour prep, essay feedback, etc. The definition of going **“above and beyond.”** (Bioethics, Fall 2022)

Milan went over arguments from papers we read very thoroughly, always addressed all student questions and encouraged student participation. He wrote helpful claims and argument maps on the board. He did **above and beyond** a GREAT job as our GSI. (Bioethics, Fall 2022)

The most distinctive aspect of Milan as a GSI was the effort he put into the course. He was incredibly helpful and the extra time he paid to individual students was an **above and beyond performance.** (Contemporary Ethical Issues, Summer 2023)

Here are the **complete** qualitative evaluations from the course I taught as a primary instructor, organized by the questions to which they respond. I’ve bolded text for emphasis, but the responses are otherwise unedited.

Do you have comments on the design of the course (including the choice of topics, readings, and assignments)?

I always enjoyed attending class because the topics were interesting and taught very organized. The slides were helpful, the examples for the concepts were perfect for comprehension.

It’s a really well-organized, logical and systematic course, probably the best course I’ve ever taken. Also topics are appealing. Readings are challenging but fun.

Milan is an amazing lecturer, and he also helped me a lot outside of the class in office hours.

Very good topics.

I liked how the quiz was based on the reading, even if most readings in itself were hard to understand.

The lectures were excellent. The way Milan explained the concepts made it easy to understand and remember. His ability to categorize complex ideas made it easy to grasp.

Shorter readings took a lot of time and some were hard to understand. I liked the reading quizzes and the outline of the exams.

All readings, assignments, quizzes helped break down all topics to make it easy to fully grasp and understand.

What was most distinctive about the way the instructor taught this course? Was anything the instructor did particularly helpful or unhelpful?

Milan always came to class ready to instruct and did so in a professional manner. He did excellent in teaching the material and is very informed about what we study. **He was always available** for questions and office hours.

Milan’s “painting” (on the blackboard) makes theory more direct and easy to understand. What’s more, the theory from original book to the slides are really clear and easy-understanding. Also office hours really solve a lot of questions I had before.

I really like the way how he teaching, he did a road map every class to help us understand how the course is going on, and explain very detail.

Help me discuss the question together and help me know how to write the paper.

The lecture slides explained the course materials thoroughly. So attending lecture and turning in was almost sufficient to doing well in the exams.

Maybe had too much on slides and didn’t finish everything (a little rushed). I did like how engaging the class has to be to understand the topics (answering questions).

Instructor was very passionate about all the subjects which was clear as he clearly made other students passionate. Helped break down difficult concepts to ensure everyone understood all material.

Any additional comments?

Milan is an amazing GSI/instructor and has been my favorite one since being at UC Berkeley. I’m confident he’ll continue to do so.

If Milan is becoming a professor at one day, I would be more than willing to applying as his students :)

Really thanks to Milan.

What was helpful is he talked about the reading during lecture so we could deeply understand or have an idea. Another thing was objection to understand the terms and situations based on information given.

10/10, would take another class with this instructor.

Complete evaluations for all of the courses I’ve TA’d are available upon request.

Summary of Quantitative Evaluations

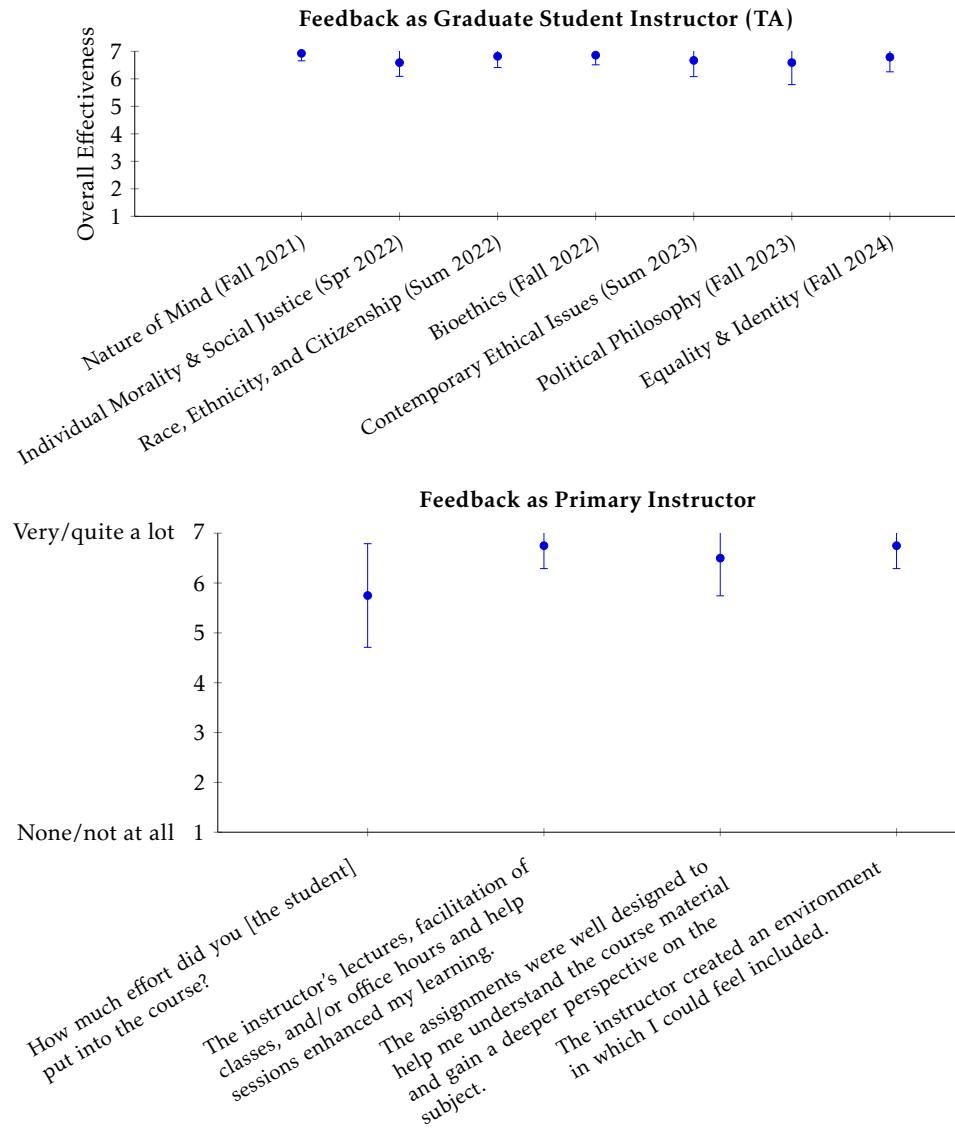


Figure 1: Quantitative Feedback. The top figure displays average (mean) student responses for the question “Overall, how effective was your Graduate Student Instructor?” The bottom figure displays average (mean) student responses for the course I taught as a primary instructor. Error bars show the sample standard deviation.

Course Descriptions

Introductory Courses

Introduction to Logic. An introduction to propositional, modal, and predicate logic, with an emphasis on connections to philosophy, mathematics, computer science, and linguistics. Students will learn how to put arguments from ordinary language into a formal language, how to construct derivations in a proof system, and compactness and completeness of propositional logic. No background in philosophy or math required.

Introduction to Moral and Political Philosophy. An introduction to central theories and issues in moral and political philosophy, with a focus on questions such as the following: What makes an action morally right? Is abortion permissible? What about breaking a promise, or killing a combatant in war? Does justice require reparations for slavery or the elimination of sex markets? Are there any objective moral requirements, or is it perfectly fine to do what is best for you, or is it all just relative? Could there be anything wrong with pursuing a genuinely meaningful relationship or life project? No background in philosophy required.

Mid-Level Undergraduate Courses

Bioethics. An introduction to clinical and population-level bioethics. We begin with questions that arise over the course of a human life, for example: What, if anything, justifies restrictions on commercial surrogacy, genetic enhancement, or abortion? What is disability? Is physician-assisted suicide permissible? We then turn to issues in population-level bioethics, for example: What is wrong with health inequality as such? How, if at all, should our healthcare system take into account people's responsibility for adverse health outcomes? In designing clinical trials and treating dangerous diseases, should we risk great harm to a few individuals, in order to save many more people from lesser but certain burdens?

Ethics and Artificial Intelligence. AI determines access to goods that we all have reason to want, given its widespread use in criminal sentencing, advertising, college admissions, grading, employment, housing, and refugee assignment. Does the use of a black-box AI violate individuals' rights to an explanation of how they are treated, or to be treated as an individual? What does it mean to align AI with human values, and how might we do that? What is discrimination, and are algorithms capable of it? How can we ensure that algorithms' decisions are fair? How might we test whether AI respects individuals' privacy? Is there a trade-off between accuracy, fairness, and privacy, and if so, how should it be struck?

Philosophy of Action. An advanced introduction to the philosophy of action, with a focus on questions such as the following: What are action, intention, and agency, and

how are they related? How can we distinguish intended and foreseen consequences of an action? Do we possess special knowledge of our intentions and actions? Why is inconsistency among intentions irrational? Are we free or responsible for our actions? What are recklessness, weakness, and compulsion, and how might we overcome them? What does it mean to share an intention to walk, dance, or cook together? When we say that a company made a decision or took an action or has responsibility, are we speaking loosely, or can we make good sense of institutions as agents, over and above the individuals who make them up?

Philosophy of Language. An introduction to the philosophy of language, with a focus on questions such as the following: How do we know what our words and sentences mean? What is meaning, and how does it relate to truth, reference, and the transmission of information? How does the meaning of a sentence depend on its parts, and on its context? What is the difference between saying something and implicating or presupposing it? What is the difference between expressing an attitude and ascribing or avowing it? How do we perform acts simply by saying words, as when taking a vow or issuing a command? Is language in some deep sense relative to a person, or to a culture? What would be lost in a society without freedom of speech?

Philosophy of Race, Ethnicity, and Citizenship. An introduction to moral, social, and political questions surrounding race, ethnicity, and citizenship, such as the following: What are race and ethnicity? What is discrimination, and why is it wrong? What is equality of opportunity, and how does racism threaten it? What is citizenship? To what extent does a state have the right to control its borders, and to assign special weight to its citizens, or to their culture? What is tolerance, and why does it matter? When does justice require civil disobedience?

Philosophy of Science. Is the aim of science truth, or explanation? Are all scientific explanations causal explanations? What is probability, and what can it explain? When does evidence confirm a hypothesis? Are theoretical entities like electrons real, or merely useful fictions? If science is objective, must it be free of value judgments? How can we distinguish science from pseudo-science? Who should receive credit for scientific discoveries? An introduction to the metaphysics and epistemology of science; philosophical problems regarding particular sciences; and ethical and social dimensions of scientific practice.

Upper-Level Courses

Causation and Discrimination. Direct racial discrimination occurs when an individual's race *causes* their worse treatment. This causal notion of discrimination is invoked by algorithmic fairness criteria, U.S. law, and correspondence studies in the social sciences. But it is difficult to spell out in a precise and plausible way, because basic features of social kinds like race and gender shine an uncomfortable light on the fundamental assumptions of causal inference. For example, a cause must be *modular*, i.e. sufficiently sep-

arable from other attributes to isolate its causal role. But gender is embedded in a nexus of social factors that resist isolated treatment: if gender is socially constructed, in what sense can it cause worse treatment? Social kinds like race and gender also pose problems for the Causal Markov condition and acyclicity, as well as the idea that causes are individuals' attributes, rather than relations between them. After an introduction to social constructionism and the basic assumptions underlying causal inference, we will discuss the problems that social kinds like race and gender pose for these assumptions, and how various thinkers have proposed to reconcile social constructionism with these assumptions. Readings from Dembroff, Haslanger, Hellman, Hu, Kohler-Hausmann, Mills, and Weinberger, among others.

Second-Personal Morality. According to contractualism, an act is right insofar as it is justifiable to each person. We will discuss the main attractions of contractualism, including its ability to place limits on interpersonal aggregation and to explain the sense in which acts can wrong particular individuals. We will also consider major criticisms of contractualism, including the worries that it cannot explain why we should save five lives rather than one; that it cannot appropriately handle cases involving risk or uncertainty; that it is too demanding; that it is uninformative or circular; and that it cannot make sense of our duties to animals and future people. Readings from Ashford, Darwall, Kamm, Kumar, Nagel, Parfit, Scanlon, Scheffler, and Wallace, among others.

Below, I've included syllabi for several courses; other syllabi are available upon request.

Syllabus Preliminaries

Instructor:	Milan Mossé	Time:	Tu/Thu 2-3:30pm
Course Website:	mmosse19.github.io	Place:	301 Phil. Hall

(To avoid repeating these preliminaries for each syllabus, I include them here.)

Grading policy: The course grade is determined by attendance and participation in lectures and sections (10%), weekly reading quizzes and discussion posts (30%), two exams (15% each), and three 2-3 page papers (10% each). I will drop your two lowest quiz scores.

Attendance: Attendance in lectures and discussion sections is required, in part because your presence and contributions contribute to the education of your peers. If you miss a class, you are responsible for the material covered; please talk to someone else in the class, come to office hours, and carefully read any handouts distributed on the material for that week.

Participation: There are lots of ways to participate in lecture, discussion sections, and office hours. Some examples: asking a clarifying question; suggesting a definition of a contested term; summarizing what someone else said in your own words, to check your understanding; (politely) raising a possible objection; expressing agreement with someone, while adding something to what they said; drawing our attention to a portion of the relevant text; and offering a relevant and concise example to help us think about a topic under discussion. Your participation grade takes into account attendance and respectful participation in lectures, sections, and office hours, as well as your discussion posts. See [here](#) for more on participation.

Email and office hours: Office hours are after class, or by appointment. Please do stop by office hours if you'd like to talk about lectures, section, drafts of papers, outside reading, or philosophy generally. If you can't attend my regular office hours (due to work, commuting to campus, a course conflict, etc.) or would prefer to discuss something privately, just let me know, and we can schedule a time. You can expect a reply to emails within two business days.

Lateness: Late papers lose a third of a letter grade, minus an additional third of a letter grade for each additional 8 hours that they are late. Everyone gets two "late days," i.e. 24 hours extensions on a paper of their choice. Extensions are always granted for emergencies—if something comes up, please don't hesitate to reach out (whenever you can).

Feedback: Your feedback is important to me. You will receive an anonymous survey after a few weeks, and another a few weeks later. I always welcome feedback over email, too.

Plagiarism and AI: [Here](#) is the university's academic honesty policy. We use plagiarism-detection software. Please don't do it—it's easy to spot and terrible for your learning and academic record. The use of generative AI in this course is forbidden.

Pronouns: As your instructor, it is my job to ensure that students are respected in the classroom. I therefore ask—and the university's Code of Conduct requires—that we all refer to each other in respectful terms, by using people's preferred names and pronouns. It is imperative that we all make a conscious effort to do this, and deliberately or repeatedly misnaming or misgendering another is unacceptable. I am responsible for ensuring that we refer to each other respectfully in the classroom, and in particular for ensuring that I refer to you using the name and pronoun you provide when introducing yourself. If you notice that I am falling short of this responsibility, or you are otherwise concerned about the atmosphere in the classroom, I welcome any feedback and encourage you to reach out to me over email or in person after class.

Accommodations: The [Disabled Students' Program](#) (DSP) organizes academic accommodations. DSP communicates directly with primary instructors, but they require students to reach out as well. Please reach out in the first or second week, so that we can make sure we have everything set up from the start of the class.

Academic resources: If you're looking for academic communities, support, resources, and events outside of this class, there are a few options:

- **Minorities and Philosophy** (MAP) has regular meetings, reading groups, and projects that you can participate in. If you're interested in joining MAP, just send an email to mapatcal@gmail.com requesting to be added to their email list.
- There is a philosophy **peer tutor program**, which I strongly encourage you to use when working on your papers for this course.
- The [Stanford Encyclopedia of Philosophy](#) (SEP) has entries that summarize the philosophical literature on various topics. [Oxford Bibliographies](#) are terser entries that collect references. Please do not cite or use Wikipedia, AI, or the dictionary to learn about philosophy—many of the entries are inaccurate, uninformative, or misleading. The SEP is much better.
- Colloquia, political theory workshops, and Cosmopolitan Philosophy talks are **all open to undergraduates**, and the last of these is geared specifically toward undergraduates.

Resources on campus for support for academic work, mental health, gender equity, disabilities, international students:

- [Counseling and Psychological Services](#)

- University Health Services
- PATH to Care Center
- Student Learning Center
- Ombuds Office for Students and Postdoctoral Appointees
- Gender Equity Resource Center
- Disabled Students' Program
- Berkeley International Office
- Centers for Educational Justice & Community Engagement

Bioethics

Description: An introduction to clinical and population-level bioethics. We begin with questions that arise over the course of a human life, for example: What, if anything, justifies restrictions on commercial surrogacy, genetic enhancement, or abortion? What is disability? Is physician-assisted suicide permissible? We then turn to issues in population-level bioethics, for example: What is wrong with health inequality as such? How, if at all, should our healthcare system take into account people's responsibility for adverse health outcomes? In designing clinical trials and treating dangerous diseases, should we risk great harm to a few individuals, in order to save many more people from lesser but certain burdens?

I. Clinical Bioethics

1. Consequentialism and Non-Consequentialism

- J. Driver, *Consequentialism* (selections)
- F. M. Kamm, "Nonconsequentialism"

2. Commercial Surrogacy

- E. Anderson, "Is Women's Labor a Commodity?"
- H. V. McLachlan and J. K. Swales, "Commerical Surrogate Motherhood and the Alleged Commodification of Children"

3. Genetic Enhancement

- M. Sandel, "The Case Against Perfection"
- F. M. Kamm, "What Is and Is Not Wrong With Enhancement? Evaluating Sandel's Views"

4. Abortion

- D. Marquis, "Why Abortion is Immoral"
- J. J. Thomson, "A Defense of Abortion"

5. Non-Identity and Disability

- D. Parfit, "The Non-Identity Problem"
- E. Barnes, "Valuing Disability, Causing Disability"
- G. Kahane and J. Savulescu, "Disability and Mere Difference"

6. Deciding Whom to Help

- P. Singer, J. McKie, H. Kuhse, and Jeff Richardson, "Double jeopardy and the use of QALYs in health care allocation"
- F. M. Kamm, "Deciding Whom to Help, Health-Adjusted Life Years and Disabilities"

7. Physician-Assisted Suicide

- J. D. Velleman, "A Right of Self-Termination?"
- Ronald Dworkin et al., "Assisted Suicide: The Philosophers' Brief" (selections)
- F. M. Kamm, "Problems With 'Assisted Suicide: The Philosopher's Brief'"

II. Population-Level Bioethics

1. Health Inequality

- D. Hausman, "What's Wrong with Health Inequalities?"
- N. Eyal, "Inequality in Political Philosophy and in Epidemiology: A Remarriage"

2. Personal and Social Responsibility for Health

- D. Wikler, "Personal and Social Responsibility for Health"
- K. Voigt, "Smoking and Social Justice"
- B. Barry, "The Cult of Personal Responsibility" (selections)

3. Paternalism, Nudges, and Incentives

- C. Sunstein and R. Thaler, "Libertarian Paternalism is not an Oxymoron"
- D. Hausman and B. Welch, "Debate: To Nudge or Not to Nudge"

4. Standards of Care

- Harvard Program on Ethical Issues in International Health Research, "The Debate Over Clinical Trials of AZT to Prevent Mother-to-Infant Transmission of HIV in Developing Nations"
- M. Angell, "The ethics of clinical research in the third world"

5. Identified and Statistical Lives

- D. Brock and D. Wikler, "Ethical Challenges in Long-Term Funding for HIV/AIDS"
- J. Frick, "Treatment vs Prevention in the Fight Against HIV/AIDS and the Problem of Identified vs Statistical Lives" in Ide

Ethics and Artificial Intelligence

Description: AI determines access to goods that we all have reason to want, given its widespread use in criminal sentencing, advertising, college admissions, grading, employment, housing, and refugee assignment. Does the use of a black-box AI violate individuals' rights to an explanation of how they are treated, or to be treated as an individual? What does it mean to align AI with human values, and how might we do that? What is discrimination, and are algorithms capable of it? How can we ensure that algorithms' decisions are fair? How might we test whether AI respects individuals' privacy? Is there a trade-off between accuracy, fairness, and privacy, and if so, how should it be struck? This course has no official prerequisites, but some experience in philosophy will be helpful.

I. Rights

This unit provides an introduction to the philosophical literature on rights that are threatened by AI. We will also discuss related problems and solutions developed by computer scientists, including explainable AI, differential privacy, and algorithmic fairness.

Week 1: The Right to Explanation

- K. Vredenburgh, "The Right to Explanation"
- S. Barocas and A. Selbst, "The Intuitive Appeal of Explainable Machines"

Week 2: The Right to Treatment as an Individual

- R. Jorgensen, "Algorithms and the Individual in Criminal Law"
- K. Piper, "The UK used a formula to predict students' scores for canceled exams"

Week 3: The Right to Privacy

- J. J. Thomson, "The Right to Privacy"
- M. Kearns and A. Roth, "Algorithmic Privacy" (*The Ethical Algorithm*, Ch. 1)

Weeks 4-5: The Right against Discrimination

- S. Moreau, "What Is Discrimination?"
- ProPublica, "Machine Bias"
- J. Kleinberg, S. Mullainathan, and M. Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores"

II. Structural Injustice, Equality, and Opportunity

AI seems to threaten several rights. But our discussion of discrimination suggested that a range of issues will be invisible to us, if we focus entirely on the rights of individuals taken in isolation, and we do not attend to the ways that AI impacts social structures. A central concern for us in this connection will be distributive and relational equality, as well as equality of opportunity. We conclude by discussing particular opportunities threatened by AI—access to information, and to the freedom to develop one's own values and make one's own choices.

1. Structural Injustice

- I. M. Young, "Political Responsibility and Structural Injustice"
- K. Creel & D. Hellman, "The Algorithmic Leviathan"

2. Distributive and Relational Equality

- D. Parfit, "Equality or Priority?"
- E. Anderson, "What is the Point of Equality?"

3. Equality of Opportunity

- T. M. Scanlon, "Procedural Fairness" & "Substantive Opportunity"
- C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel, "Fairness Through Awareness" (selections)

4. Opportunities Information, Values, and Choice

- The New York Times, "They Asked an A.I. Chatbot Questions. The Answers Sent Them Spiraling."
- C. Thi Nguyen, "Value Capture"
- C. Castro & A. K. Pham, "Is the Attention Economy Noxious?"

III. Alignment

This unit provides an introduction to the conceptual foundations, failure modes, and current methods for AI alignment.

1. Introducing Alignment

- AISafety.info, "The alignment problem"
- D. Parfit, "The Non-Identity Problem"

2. Time and Risk

- AI Alignment Forum, “AGI Safety from First Principles”
- J. Frick, “Contractualism and Social Risk”

3. Methods

- Bai et al., “Constitutional AI”
- Amazon Web Services, “What is RLHF?”

Introduction to Moral and Political Philosophy

Description: An introduction to central theories and issues in moral and political philosophy, with a focus on questions such as the following: What makes an action morally right? Is abortion permissible? What about breaking a promise, or killing a combatant in war? Does justice require reparations for slavery or the elimination of sex markets? Are there any objective moral requirements, or is it perfectly fine to do what is best for you, or is it all just relative? Could there be anything wrong with pursuing a genuinely meaningful relationship or life project? No background in philosophy required.

I. Ethical Theory and Problems

1. Utilitarianism and Beneficence

- P. Singer, "Famine, Affluence, and Morality"
- H. Sidgwick, *The Methods of Ethics* (Book IV, Ch. 1: The Meaning of Utilitarianism)

2. Rights Theory and Abortion

- F. M. Kamm, *Morality, Mortality* (pp. 5-10)
- J. J. Thomson, "A Defense of Abortion"

3. Kantianism and Lying

- I. Kant, *Groundwork of the Metaphysics of Morals*, Cambridge ed. (pp. 9-18, 26-34, 40-42)

4. Virtue Ethics and Abortion

- G. E. M. Anscombe, "Modern Moral Philosophy"
- R. Hursthouse, "Virtue Theory and Abortion"

5. Contractualism and War

- T. Nagel, "War and Massacre"
- T. M. Scanlon, *What We Owe to Each Other* (Ch. 5, selections highlighted in pdf)

II. Political Philosophy and Problems

1. Liberalism and the Family

- J. Rawls, *Theory of Justice* (selections)
- V. Munoz-Dardé, "Is the family to be abolished then?"

2. Egalitarianism and Artificial Intelligence

- D. Parfit, "Equality or Priority?" (pp. 81-85, 97-99)
- E. Anderson, "What is the Point of Equality?" (selections)
- J. Angwin, J. Larson, S. Mattu and L. Kirchner, *Machine Bias*

3. Libertarianism and Reparations

- R. Nozick, *Anarchy, State, Utopia* (selections)
- B. Boxill, "A Lockean Argument for Black Reparations"

4. Solidarity and Markets

- M. Zhao, "Solidarity, Fate-Sharing, and Community"
- W. Hussain, "Pitting People Against Each Other"

5. Political Obligation

- S. Scheffler, "Membership and Political Obligation"

III. Metaethics and Moral Psychology

1. Justice and Egoism

- Plato, *The Republic*, Books I-II (327a - 369a)
- J. Rachels, "Two Arguments against Ethical Egoism"

2. Relativism and Anti-Realism

- H. Gensler and M. G. Tokmenko, "Against Cultural Relativism"
- S. Street, "Does Anything Really Matter, or Did We Just Evolve to Think So?"

3. Luck and Meaning

- *Complete Works of Zhuangzi*, trans. B. Watson (§§15, 18, 26)
- B. Williams, “Moral Luck” (selections)
- S. Wolf, “The Meanings of Lives”

4. Freedom and Transformation

- H. Frankfurt, “Freedom of the Will and the Concept of a Person”
- L. A. Paul, “What You Can’t Expect When You’re Expecting”

Film Screening: *Ikiru* (A. Kurosawa)

Philosophy of Science

Description: Is the aim of science truth, or explanation? Are all scientific explanations causal explanations? What is probability, and what can it explain? When does evidence confirm a hypothesis? Are theoretical entities like electrons real, or merely useful fictions? If science is objective, must it be free of value judgments? How can we distinguish science from pseudo-science? Who should receive credit for scientific discoveries? An introduction to the metaphysics and epistemology of science; philosophical problems regarding particular sciences; and ethical and social dimensions of scientific practice.

I. Explanation

1. Scientific Explanation

- M. Friedman, "Explanation and Scientific Understanding"
- W. C. Salmon, *Four Decades of Scientific Explanation* (pp. 46–50)

2. Mathematical Explanation

- I. Lakatos, *Proofs and Refutations* (selections)
- P. Mancosu, "Mathematical Explanation"

3. Causal Explanation

- J. L. Mackie, "Causes and Conditions" (selections)
- D. Lewis, "Causation"
- N. Cartwright, "Causal laws and Effective Strategies"

II. Probability, Confirmation, and Bayesian epistemology

1. Probability

- R. Briggs, "The Metaphysics of Chance"
- B. Loewer, "David Lewis's Humean Theory of Objective Chance"

2. Confirmation

- N. Goodman, *Fact, Fiction, and Forecast* (selections)
- C. G. Hempel, "Studies in the logic of confirmation"

3. Bayesian Epistemology

- M. Strevens, "Notes on Bayesian Confirmation Theory"

III. Realism and Sciences

1. Physics

- A. Franklin, "Are There Really Electrons? Experiment and Reality"
- B. van Fraassen, "Arguments Concerning Scientific Realism"

2. Biology

- I. Hacking, "Do We See through a Microscope?"
- K. Okruhlik, "Gender and the Biological Sciences"

3. Economics

- C. Bicchieri, "Rationality and Indeterminacy"
- U. Mäki, "Realistic Realism about Unrealistic Models"

IV. Science and Society

1. Values and Science

- H. Longino, "Values and Objectivity"
- E. Anderson, "Uses of Value Judgments in Science"

2. Pseudo-Science

- C. McKinnon, "Should We Tolerate Climate Change Denial?"
- P. R. Thagard, "Why Astrology is a Pseudoscience"

3. Cognitive Labor and Credit

- P. Kitcher, "The division of cognitive labor"
- M. Strevens, "The role of the priority rule in science"