

A GUIDE TO WRITING PHILOSOPHY PAPERS AT CALTECH

This resource provides detailed guidelines for writing philosophy papers in Caltech courses, along with specific examples taken from prizewinning essays written by Caltech students. It was created through a collaborative effort between the [Hixon Writing Center](#) and the philosophy faculty members at Caltech.

The goal of philosophical writing at Caltech is for students to develop the ability to construct a clear, well-supported argument that states its assumptions and considers its weaknesses. This involves not only presenting a compelling case but also thoroughly laying out the possibilities of how a philosophical question might be addressed, recognizing the pros and cons of each option. In writing philosophy papers, students learn to develop and present the best *objective reasons* for a point of view rather than merely expressing a personal opinion.

TYPES OF PAPER ASSIGNMENTS

Understanding an essay prompt is the first step toward writing a successful philosophy paper. Read the prompt several times to ensure a complete grasp of its key questions and expectations. Identify the main philosophical ideas you'll be asked to work with and be sure that you understand whether you are being asked to describe another writer's ideas, critique their ideas, and/or offer your own original ideas that build on their work.

Philosophy assignments can vary widely, with each type demanding a specific approach. Below are some common types of paper assignments you might encounter in your Caltech philosophy courses, followed by example prompts:

- **Provide a critical assessment of a philosopher's argument:**

"Explain and critically evaluate Descartes' argument that God exists and is not a deceiver."

- **Assess a debate between two or more philosophers to determine who has the stronger position:**

"Loewer and Maudlin present opposing views on what laws of nature are. Which view do you think is correct, and why?"

- **Develop and defend a specific interpretation of a philosophical text (typically one where the interpretation is controversial):**

"Descartes tries to achieve knowledge with certainty. What, precisely, is his goal? Is it his own subjective psychological certainty? Or does he want objectively sound

arguments? Is he only trying to convince himself or is he trying to convince others as well? Provide textual evidence in support of your interpretation.”

- **Apply ideas from a philosopher to a novel situation:**

“From a utilitarian perspective, how much effort should we be putting into colonizing other planets?”

- **Develop a philosopher’s ideas in an original way:**

“Parfit poses the question: ‘What makes a person at two different times one and the same person?’ Explain the question and argue for a particular answer, comparing your answer to those proposed in the readings by Parfit and Lewis. In your essay, you may want to use examples involving teleportation and/or parallel universes.”

- **Discuss how a philosopher’s views might need to be modified in light of 21st-century scientific knowledge:**

“Explain Locke’s distinction between primary and secondary qualities. Which qualities does he classify as primary and which as secondary? If Locke had 21st-century scientific knowledge, would he change his mind about which qualities are primary and which are secondary?”

The audience for your essay will depend on the level of the course you are taking and the specifics of the assignment. Some professors might ask you to write for other students in the class or an educated adult audience familiar with basic philosophical reasoning. Other professors might ask you to write for Caltech students who are *not* in the course, meaning you should explain the philosophical ideas you discuss and not assume that the reader is already familiar with them. Carefully read the assignment and/or provided course materials to understand audience expectations.

PRESENTING THE GOAL OF YOUR PAPER

Toward the beginning of your essay, clearly and concisely state the goal of your paper to tell your reader what you will accomplish. The goal of your essay should be modest so that it can be explained and supported in a relatively short paper. Try to formulate goals that are appropriate to your knowledge of the subject and the length of the paper.

Early in your essay, provide enough context to situate your reader. While it may be tempting to weave an argument into an extended broader narrative or historical context, as some professional philosophers do, directness is typically more appropriate in writing for course assignments. You will need space to go into depth on the merits of a particular argument, and the broader context can take that space away. Your paper should, however, provide some exposition that summarizes the ideas you will work with. See the section on Engaging with Other Philosophical Works below for more information.

Example

"The key to short-term happiness, then, is to expect good things, but not to expect great things. This essay will argue that low but positive expectations are vital to having a good mood in the short term. I will also illustrate how lifestyle changes incorporating this philosophy not only contribute to short-term happiness but also contribute to long-term satisfaction."

From McClure Prize winner Lark Mendoza's ["Good Expectations: The Paradox of Expecting Great Things"](#)

"In this paper, I will argue that an AI system should have full moral status if and only if it possesses some degree of sentience—the ability to hold beliefs, have experiences, and/or feel emotions."

From Hixon Prize winner Elsa Palumbo's ["Utilitarian Calculations and the Moral Status of Strong AI"](#)

DEFINING AND USING PHILOSOPHICAL TERMS

Define all philosophical terms that you use. While many terms might sound commonplace, they often have special meanings in philosophy. For instance, terms like "foundationalism" or "dualism" might be understood differently in different contexts. When such specialized terms appear in your paper, you must offer clear definitions.

Resist the urge to quote the dictionary. If you are using a word in its ordinary English sense, you can assume the reader knows it. If you are not using it in its ordinary sense (i.e., if it has a special philosophical meaning), the dictionary will not be helpful here. Instead, explain its philosophical meaning to ensure your reader understands your perspective. And if you are using a term in the way a particular philosopher used it or borrowing their definition, be sure to cite that author.

Jim Pryor's [Philosophical Glossary for Beginners](#) is a good place to find definitions of commonly used philosophical terms. Note that seemingly straightforward terms like ["logical"](#) and ["valid"](#) have a particular meaning for philosophers and should be used correctly.

Once you've provided definitions of your terms, the next challenge is maintaining that definition throughout your essay. Having set a term's definition, it's important to stick to that meaning consistently. Avoid vacillating between different definitions that might confuse your reader. Do not use multiple words as synonyms for the same concept.

Example

"Functionalism is the view that there is only one kind of substance, physical substance, and that the nature of a mental state is independent of its physical realization. This means that a mental state is solely defined by the causal or functional role it plays with regard to the external world, including other mental states. Thus, the key points of functionalism are that: 1) There are only physical substances, no separate mental substances 2) There could be drastically physically different systems that are still minds and have mental states, provided the states of that system play the same functional roles as those of the mental states of the brain."

From Hixon Prize winner Vansh Tibrewal's ["The Chinese Room Argument"](#)

PROVIDING A ROADMAP

At the outset of your essay, in the first or second paragraph, it is often helpful to provide your reader with a roadmap, a brief overview of the path your argument will take.

Your roadmap should lay out a clear plan of argumentation for your reader. It serves as an outline of the strategy you will use to support your argument. In a philosophy paper, no part of the argument should come as a surprise. If a reader reaches the end of your paper and is surprised by your conclusion, it suggests the argument may not have been sufficiently clear. During the revision process, be sure to confirm that you've stuck to your roadmap throughout the essay.

A roadmap serves two purposes. It sets expectations and provides a structure that the reader can refer to if they need to recall the paper's trajectory. The roadmap can take a first-person approach. Signposting phrases like "First, I will argue" and "Then, I will discuss" are the foundation of a good roadmap.

Example

"In this paper, I will argue that an AI system should have full moral status if and only if it possesses some degree of sentience—the ability to hold beliefs, have experiences, and/or feel emotions. In doing so, I will start by anticipating and responding to four possible objections to giving sentient AI systems full moral status, focusing especially on issues concerning in-group favoritism, autonomy, uncertainty about AI's thinking ability, and the intensity of humans' pleasure and pain. Notably, unlike writers Bostrom and Yudkowsky, I do not claim that sapience—self-awareness and other qualities associated with higher intelligence—has a role in determining whether humans and an AI system should be given equal moral weighting, and I will defend this claim partly through arguments inspired by Peter Singer's book *Animal Liberation*. Unlike Singer, however, I will argue that sentience is a better criterion than a capacity for suffering, since the former leaves open the option of assigning full moral status to emotionless Strong AI systems with a rich inner life, like the android Data in *Star Trek*."

From Hixon Prize winner Elsa Palumbo's ["Utilitarian Calculations and the Moral Status of Strong AI"](#)

SIGNPOSTING, TOPIC SENTENCES, AND WRITING IN THE FIRST PERSON

Signposting, in the context of writing, refers to the use of specific phrases or statements that guide the reader through the structure and progression of your arguments. These phrases act like road signs, indicating where the argument has been, where it is currently, and where it is heading. This is particularly important in philosophy papers, where the progression of thought and rigorous argumentation must be easy for the reader to follow.

We've already shown how you can use signposting at the beginning of your paper to provide readers with a roadmap. A good roadmap helps orient readers and sets expectations for the structure of your paper.

Similarly, topic sentences at the start of paragraphs can act as signposts. They can either tell the reader what the paragraph will discuss or make a claim that the paragraph will then support. Examples include phrases like: “I will now delve into,” “I will now explain,” or “An important consideration here is.” An essay that provides a good roadmap and effective signposts stands a better chance of maintaining *coherence*, a concept we explain toward the end of this resource.

While you may have been taught at some point *not* to use the first person in academic writing, philosophical writing does not abide by this rule. Philosophers frequently make use of the first person to make their arguments clear. Writing in the first person can make it clear when you are setting forth your position rather than describing someone else’s idea. A phrase like “I will argue that” makes it evident that what follows is *your* stance rather than that of another thinker and removes any ambiguity about your position.

While personal opinions and preferences can motivate your arguments, it’s important to write objectively when using the first person, supporting your claims with arguments. For this reason, some professors may ask you not to use phrases like “In my opinion” or “I ultimately prefer.” The strength and logic of your arguments lend credence to your conclusions, not personal preference.

Example

“Now that I’ve explained why I think it is problematic for us to deny certain AI systems full moral status, I want to address four potential objections to my position.”

From Hixon Prize winner Elsa Palumbo’s [“Utilitarian Calculations and the Moral Status of Strong AI”](#)

“Having put to bed the issue of whether strong AI proponents are dualists or functionalists, I will now explain the Chinese Room Argument as laid out by Searle.”

From Hixon Prize winner Vansh Tibrewal’s [“The Chinese Room Argument”](#)

SUPPORTING YOUR ARGUMENT

Most of your essay should consist of reasoning that supports your arguments. It’s not enough to simply state your view; you must also persuade the reader that it is correct. The specific kind of support you offer will depend, in part, on what type of essay you are writing. The goal of a philosophy essay is not necessarily to pile up separate pieces of supporting evidence. Instead, the evidence should be carefully selected to build a specific argument. For example, your essay might focus on defending or refuting one particular supposedly definitive argument for a particular philosophical position. And while empirical evidence from other disciplines, such as neuroscience or physics, may be useful in supporting certain arguments, you must still explain why this evidence is pertinent and what philosophical conclusions can be drawn from it.

Examples or hypothetical scenarios might be particularly useful when used as part of a philosophical argument, objection, or reply. Examples can also be helpful when explaining philosophical ideas and technical terms. A good example can make an abstract concept clearer.

It can be challenging to find clear examples that illustrate a philosophical point. For this reason, finding one clear example that you can use with slight modifications throughout the essay as a

running example may be useful. Examples should be chosen and presented with an eye towards efficiency, eschewing extraneous details. Aim to isolate and articulate the point you are trying to make while minimizing irrelevant complications.

Example

“Finally, consider Singer’s argument that sentience alone is not sufficient for full moral status, because beings need emotions and/or a capacity for pain in order to suffer. Here’s my response: Following Singer would mean excluding from our calculations individuals like *Star Trek*’s android Data who, despite having no emotions and being incapable of experiencing pain, still had valuable interpersonal friendships, a unique perception of reality, an acute sense of self, and life-long goals of self-improvement—as well as a strong sense of self-preservation, as became clear in the episode *The Measure of a Man*, when an AI researcher threatened to dismantle him for the sake of science. In particular, Data is sentient, if we continue to use my definition of sentience as “the ability to hold beliefs, have experiences, and/or feel emotions.” In fact, although Data does not feel emotions, he certainly holds strong convictions about what is right and wrong, how the world works, and what his role in it is. Moreover, he observes and interprets experiences through his own lens, one colored by his individuality, making it clear that Data does still fit under my definition of sentience on two grounds: he does (1) hold beliefs and (2) have experiences of his own. All things considered, it seems immoral to exclude sentient but emotionless individuals like Data from our moral calculations.”

From Hixon Prize winner Elsa Palumbo’s [“Utilitarian Calculations and the Moral Status of Strong AI”](#)

ENGAGING WITH OTHER PHILOSOPHICAL WORKS

Almost all modern philosophical writing includes substantial, explicit engagement with the writing and ideas of other philosophers, and your course assignments may also require this. The best philosophy essays give careful and detailed summaries of relevant points from the readings while remaining concise.

Philosophical writing tends to use direct quotations strategically, often using paraphrases so that an author can present their particular understanding of a text. The goal is not simply to rephrase an author’s ideas but to put them into your own words to show that you fully understand them. An effective paraphrase will generously and accurately represent another writer’s ideas.

When a specific passage from a text critically underpins your interpretation of a philosopher’s views, directly quoting this passage can be useful. Ensure you precisely indicate the source of the passage by citing it. Nevertheless, the use of direct quotations should be the exception, not the rule. Typically, quoting a few pivotal sentences is sufficient. However, in cases where a text’s wording could lead to multiple interpretations or where specific phrasing is crucial for your analysis, a brief, direct quote can be particularly insightful. Quotations should never replace your own explanation. Even when quoting an author, you should elucidate the meaning of the quotation in your own words. See the example below.

When working with sources, you should avoid summarizing parts of a philosopher’s work that are irrelevant to the assignment, but you should also be careful to present those parts that will be the focus of later critique. For example, if you think that Descartes’ argument for the

existence of God fails at a particular step, the exposition of the argument should make that step explicit. While the point of a philosophy essay is not to demonstrate everything you know about a particular philosopher or idea, your essay *should* include some exposition that summarizes the positions and arguments that you will engage with. For this reason, some professors might ask you to imagine writing for other students who have not taken the course or done the readings. Providing such background will help you make your own philosophical contribution, which is the ultimate goal of a paper.

When you present someone else's argument, be sure to present the strongest possible formulation of that argument. This should be done in all academic writing, but it is particularly important in philosophical writing, which calls for a high degree of precision. Many philosophy essays fail specifically because they mischaracterize the work of other philosophers, leading to a straw man argument.

Example

"Consider first Camus's description of how we might find absurdity by simply observing nature: '[S]trangeness creeps in... these hills, the softness of the sky, the outline of these trees at this very minute lose the illusory meaning with which we had clothed them... The world evades us because it becomes itself again' (14). This passage is reminiscent of Sartre and his famous work *Nausea*; Camus urges us to see that we have constructed meaning from a world that is stark and raw in its absolute form. In specific moments, when we see the world for what it is, its imagined meaning is only 'illusory' and slips from our grasp."

From McClure Prize winner Kshitij Grover's "[Living with Sisyphus](#)"

CITING OTHER PHILOSOPHICAL WORKS

Professional philosophers use a variety of citation systems, such as Chicago or APA, depending on where they are publishing their work. It is important to determine whether your course sets specific standards for citation style. Consult the Purdue OWL's [Research and Citation Resources](#) for information on specific citation styles.

No matter what citation style you are asked to use, it is important to provide a reference whenever you present another author's ideas. Likewise, you should provide a reference the first time you mention a specific work. Both of these situations are shown in the example below, which uses APA Style. Failing to cite sources properly could constitute [plagiarism](#).

Example

"First, I will discuss the arguments Searle is responding to. In his *Minds, Brains, and Programs*, Searle focuses on the work of Roger Schank and Abelson at Yale (Searle 1997, 183). Schank aims to produce a machine that can understand stories, similar to humans. He asserts that this ability can be measured through the capability to answer questions about the story. He claims a machine can exist that can be given a story in an appropriate representation and be able to answer questions about it, drawing on the information it has been previously fed, in a manner similar to a human would. Schank makes the following key claims: 1) the machine in question understands the story 2) the functioning of the machine explains how human understanding works. Searle will

challenge both of these claims. Further, although Searle does not directly mention Turing, his arguments pose a challenge to the validity of the Turing Test. The Turing Test, proposed by Alan Turing in 1950, is a means by which to test whether a machine can think. More precisely, Turing claims that it is vague and pointless to ask whether a machine *thinks*, and instead states that it is sufficient to just test whether the machine acts like it thinks (Turing 1997, 29-30). Searle's thought experiment challenges the validity of the Turing Test by proposing a hypothetical whereby the machine would pass the Turing Test and yet seem to not truly *think* or *understand*, at least as we currently conceive of the terms intuitively."

From Hixon Prize winner [Vansh Tibrewal's "The Chinese Room Argument"](#)

ANTICIPATING AND ADDRESSING OBJECTIONS

In many types of philosophy papers that you might write at Caltech, it will be useful to anticipate and address possible objections to your argument by offering a counterargument. Effectively dealing with objections enhances the robustness of your main argument and adds depth to your reasoning. It is much more effective to confront possible objections yourself than to leave them for your reader to uncover. Given the constraints of an undergraduate philosophy paper, addressing every possible objection to your argument is unnecessary. Focus instead on the most significant objections.

Example

"A second counterargument is inspired by John Searle's Chinese Room. This thought experiment goes as follows: imagine Searle alone in a room with detailed instructions on what combination of Chinese characters should be used to respond to any sequence of Chinese characters (a question) that gets slipped under the door. Provided that the instructions (programming) are detailed enough, Searle (an AI system) may, through the mindless manipulation of symbols, convince native Chinese speakers outside the room that he actually understands Chinese (pass the Turing test), even though he does not. By similar reasoning, we can never be sure that an AI system is truly sapient or sentient, and not merely a very skillfully programmed fraud. So, the argument goes, even if we agreed that an AI system known to possess some particular combination of sapience and/or sentience should be given full moral status, we could never be certain that the conditions are satisfied.

As with many skeptical arguments, this one can be de-thorned fairly easily. The issue here is confusing certainty with knowledge. All that is required for knowledge is a justified true belief, and there are ways to obtain sufficient justification. Even if the Chinese Room argument has convinced us that direct observation of an AI system is not sufficient, there's another strategy: look closely at the system's programming and check how directly output follows from input. The greater the complexity of the process that arrives at an answer, the greater the chances that the system is not merely (1) getting a symbol that, to it, is meaningless, (2) matching it with instructions, and (3) returning another symbol. Hence, even though this method would not be foolproof, greater complexity also gives us stronger evidence that true thinking/feeling is taking place."

From Hixon Prize winner Elsa Palumbo's ["Utilitarian Calculations and the Moral Status of Strong AI"](#)

WRITING THE CONCLUSION

The concluding paragraph of a philosophy paper usually consists of a straightforward restatement of the essay's content. Here, you should give your reader a concise summary of your main points. See the example below.

You could encounter an assignment that asks you to accomplish other tasks in your conclusion, such as pointing to future work that could be done on the topic or elucidating the current work's limiting factors. If that's the case, follow those instructions.

Example

"In order to arrive at the public policy recommendations, I first established that quantitative hedonism is superior to qualitative hedonism in aligning our intuitions about life on the *Axiom* with a theory of well-being. Then, I used quantitative hedonism to examine Cohen's features of "good jobs" and produced two conditions for justifying automation: 1) the worker whose job was automated must be able to obtain more utility from a "safety net" than they did while working, and 2) the automated job must yield more total utility than it did prior to being automated. Lastly, I proposed a three-pronged approach to public policy to regulate the automation of labor, and I applied these policy recommendations back to the world of the *Axiom*. Overall, I believe that public policy that regulates the automation of labor at a steady pace will allow us to reap the same benefits as those on the *Axiom*, without losing elements of our society that provide us large amounts of pleasure."

From Hixon Prize winner Avirath Sundaresan's ["WALL-E Wisdom: Lessons for Public Policy in the Age of AI"](#)

KEY FEATURES OF CLEAR ACADEMIC WRITING

Philosophy often deals with abstract, complex subjects, so philosophical writing should strive for clarity. Clear writing exhibits a high degree of **coherence**, **cohesion**, and **conciseness**. The characteristics listed in this section are general principles for academic writing, but they are especially important for philosophical writing.

An essay demonstrating **coherence** focuses on a specific topic and purpose rather than jumping around. To judge coherence, we can look at the paper as a whole and at the paragraph level. Do all of the paragraphs contribute to the essay's main argument, and do they appear in a logical order? As we explained earlier in the handout, philosophy papers use roadmaps and signposting to help maintain coherence and guide the reader.

In cohesive writing, sentences connect to one another so that the reader can easily follow the writer's ideas from sentence to sentence. You can ensure that your writing is cohesive by moving from old information to new information in individual sentences and judiciously using transition words and phrases between sentences like "therefore," "however," "furthermore," and "in contrast."

When writing is **concise**, it does not contain unnecessary words and phrases. Sentence structure tends toward simplicity (e.g., subjects preceding verbs; avoiding linking many phrases and clauses together), with longer or complex sentences being used sparingly. Philosophers

especially value simple, straightforward language. As you revise, focus on simplifying sentences and eliminating unnecessary words.

Note that these features only characterize the writing itself. An excellent essay must demonstrate a deep understanding of the readings and present strong original ideas and arguments.

ADDITIONAL RESOURCES

Prize-Winning Philosophy Papers by Caltech Writers

- Natasha Hong, [“Desert Under Unequal Conditions: Meritocracy, Luck, and Moral Worth”](#) (2026)
- Mariana Taveira, [“From Carbon to Silicon: A Philosophical Case for Surviving Through Mind Uploading”](#) (2025)
- Edward Speer, [“The Good Life+: Well-Being in Virtual Worlds”](#) (2025)
- Sarah Yuan Ni Liaw, [“The Paradox of Authentic Happiness and Existential Suffering”](#) (2024)
- Vansh Tibrewal, [“The Chinese Room Argument”](#) (2023)
- Lark Mendoza, [“Good Expectations: The Paradox of Expecting Great Things”](#) (2023)
- Elsa Palumbo, [“Happiness is Hedonic: In Defense of a Subjective Account of Well-Being”](#) (2022)
- Ryan White, [“Fairness, Morality, and Pursuing an Ideal System of Distributive Justice”](#) (2021)
- Elsa Palumbo, [“Utilitarian Calculations and the Moral Status of Strong AI”](#) (2020)
- Lucca de Mello, [“What Criteria Should Be Used to Determine Whether an Automated Decision Procedure Is Fair?”](#) (2020)
- Avirath Sundaresan, [“WALL-E Wisdom: Lessons for Public Policy in the Age of AI”](#) (2022)
- Galilea von Ruden, [“Why the Turing Test Revised Is Still the Turing Test”](#) (2018)
- Kshitij Grover, [“Living with Sisyphus”](#) (2017)
- William Hoza, [“A Can of Worms”](#) (2016)

Additional prize-winning essays by Caltech writers can be found [here](#).

Hixon Writing Center Resources

- [Working with Sources and Writing About Philosophy](#)
- [Avoiding Plagiarism: Guidelines and Expectations for Writing in HSS](#)
- [Writing and Procrastination](#)
- [Making the Time to Write \(And Actually Writing\)](#)
- [Generating a Draft without Getting Blocked](#)
- [Revising with Reverse Outlines](#)

Recommended Philosophy Writing Guides

- Jim Pryor’s [Guidelines on Writing a Philosophy Paper](#) is a particularly valuable resource.
- See also Pryor’s [Philosophical Terms and Methods](#), which provides detailed information on the nature of philosophical arguments and how to make them.

WORKS CONSULTED

In addition to the Pryor guides noted above, the authors of this handout consulted several other resources during the writing process:

- The Harvard Writing Center's [A Brief Guide to Writing the Philosophy Paper](#)
- [The Pink Guide to Philosophy](#)
- Justin Weinberg's [Writing a Good Philosophy Paper](#)
- The University of Michigan's [Tips for Writing Philosophy Papers](#)

This document is licensed under Attribution-NonCommercial 4.0 International. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

