1 When is nominalization good?

We cannot write without nominalization. The trick is to know which to keep and which to revise.

1. A nominalization that is a short subject:
   a. These results show great promise. \textit{(results)}
      (The above sentence is better than \textit{These results are promising} because \textit{show} is a stronger verb than \textit{are}.)
   b. This operation is polymorphic. \textit{(operation)}

2. A short nominalization that replaces an awkward \textit{The fact that}:
   a. The fact that the program performs well speaks to its promise.
   b. The good performance of the program speaks to its promise.

3. A nominalization names what would be the object of the verb:
   a. The server responded to what the client requested.
   b. The server responded to the client’s request.

4. A nominalization that is a character:

   My research area is high-performance computing.

2 Identify characters

   a. The CPU computed the result of integrating the function that characterized power consumption.
   b. The CPU performed a computation to determine the integral of the function for the characterization of power consumption.
   c. The computation was performed by the CPU for the determination of the integral of the function that gives the characterization of the power consumption.

(a) and (b) are better than (c) because they have characters as subjects. \textit{(CPU, function)}

Williams says that readers want action in verbs, but even more they want characters as their subjects. We give readers a big problem when, for no good reason, we do not name characters in subjects, or worse, delete them entirely.

Williams proposes three steps to diagnose a sentence:
1. Underline the first seven or eight words;

2. Find the main characters.

3. Skim the passage for actions involving those characters, particularly action buried in nominalizations. Ask *Who is doing what?*

To revise, reassemble those new subjects and verbs into a sentence, using conjunctions such as *if, although, because, when, how, and why:*

a. Entry of data into the processor chip causes the direction of the data into the cache and it staying there until ejection.

b. When data enters the processor chip, it is directed into the cache and stays there until it is ejected.

### 3 Passive and active voice

Passive voice is OK, as long as the characters are serving as subjects. It is fine to use *I, we, or the authors.* But usually *we* means the authors and readers together. And the second person *you* is also good because it draws the readers in, making them feel engaged.

a. This chapter will familiarize you with the framework we shall use throughout the book to think about the design and analysis of algorithms.

b. I/The authors discovered this algorithm during the course of my/our work on using column-sort as the basis for out-of-core sorting.

c. If you have been introduced to any of these languages, you should have little trouble understanding the algorithms.

### 3.1 What is passive and active voice?

“The terms *active* and *passive,* however, are ambiguous, because they can refer not only to those two grammatical constructions but also to how a sentence makes you feel. We call a sentence *passive* if it feels flat, regardless of whether its verb is grammatically in the passive voice.” (Williams) For example:

a. The program will run fast if it creates few cache misses.

b. The speed of the program depends on creation of few cache misses.

Both of above sentences are in active voice. But the first one feels more active and more concrete, because it has good characters (program and cache misses) as characters. The second one feels passive because the characters are buried and the verb is not the action.
3.2 How to choose between passive and active?

And Williams suggests we answer three questions to choose between active and passive:

1. Must your readers know who is responsible for the action?

2. Would the active or passive verb help your readers move more smoothly from one sentence to the next? It’s good to start a sentence with the familiar and introduce new notions later in the sentence.

   a. Although software caching was a common technique by the early 1960s, memory access times decreased in large part due to the advent of caches in hardware. The first hardware caches [familiar] were designed in the mid 1960s at IBM [new].
   b. Although software caching was a common technique by the early 1960s, memory access times decreased in large part due to the advent of caches in hardware. IBM [new] designed the first hardware caches [familiar] in the mid 1960s.

   The first one is better because it starts with the familiar (caches, what the paragraph is talking about) and ends up with the new information (IBM). Make what the sentence focuses on the subject of the sentence.

3. Would the active or passive give readers a more consistent and appropriate point of view?

   a. By early 1945, the Allies had essentially defeated Germany; all that remained was a bloody climax. American, French, British, and Russian forces had breached its borders and were bombing it around the clock. But they had not yet so devastated Germany as to destroy its ability to resist.
   b. By early 1945, Germany had essentially been defeated; all that remained was a bloody climax. Its borders had been breached by American, French, British, and Russian forces, and it was being bombed around the clock. It had not been so devastated, however, that it could not resist.

   Both of those two paragraphs are OK. The first one explains the history from the point view of the Allies, while the second one of German. Pick a point of view and stick to it. Don’t switch one for another back and forth. It also makes it hard to figure out whom the pronoun is referring to.

   To sum up, both nominalization and passive voice are fine. The key is to understand what you are doing.

3.3 Passive, Characters, and Metadiscourse

In scientific writing, verbs fall into two groups:

1. Some refer to research activities: examine, observe, measure, record, use, study. Those verbs often appear in the passive voice.

2. Others refer not to the subject matter or the research, but to the writer’s own writing and thinking: cite, show, inquire. Those verbs are often active and in the first person. They are examples of metadiscourse.

   Metadiscourse is language that refers not to the substance of your ideas, but to yourself, your reader, or your writing:
1. Your thinking and act of writing: explain, show, argue, claim, deny, suggest, contrast, add, expand, summarize, prove, see, cite, require.

2. Your readers’ actions: consider now, as you can recall. They are used to direct the readers.

3. The logic and form of what you have written: first (not firstly), second (not secondly); to begin; therefore, however, consequently; seems; undoubtedly.

Passive sentences can create dangling modifiers — the introductory phrase has an implied subject that differs from the explicit subject in the following or preceding clause. For example:

a. To determine which program was faster, we ran them on a cluster.

b. To determine which program was faster, they were run on a cluster.

4 Cohesion and Coherence

4.1 Cohesion

“Sentences are cohesive when the last few words of one set up information that appears in the first few words of the next. That’s what gives us our experience of flow.”

a. In order to reduce latency in accessing data, computer architects started designing caches in the 1960s. Fast, associative memory forms caches. Because such memory is relatively expensive, caches are much smaller than main memory.

b. In order to reduce latency in accessing data, computer architects started designing caches in the 1960s. A cache is typically made from a fast, associative memory. Because such memory is relatively expensive, caches are much smaller than main memory.

The first paragraph is not cohesive because the first sentence ends with “cache” while the second starts with “associative memory.” It doesn’t flow smoothly.

It is fine to use passive voice in order to make the paragraph cohesive.

Usage: The whole comprises the parts.

Sidetrack: When to put commas within multiple adjectives?

Put a comma in between if the adjectives are of equal significance. Rule of thumb: if adding and in between is OK, then put a comma (fast, associative memory); if not, do not put a comma (a wiry old carpenter).

4.2 Coherence

Coherence is the sense of whole. Do all the sentences fit together? It is possible that the sentences are cohesive, but the whole passage is incoherent.

To make a passage coherent, use a topic sentence (even a pseudo-topic, “The following paragraph is about”), make everything pertinent to the topic, and make your conclusion in the last sentence.
5 Emphasis

a. Average-case analysis and experimental results of the algorithms for sorting and searching on random data to determine which algorithms are best in practice are the subjects of this section.

b. In this section, we analyze sorting and searching algorithms on random data, examining average cases and experimental results to determine which algorithms are best in practice.

The second one is better because it leaves the important point to the end.

5.1 Complex grammar

a. Natarajan’s assertion that columnsort takes longer than dsort because columnsort makes one additional pass appears in the last section of her paper.

b. In the last section of her paper, Natarajan asserts that columnsort takes longer than dsort because columnsort makes one additional pass.

“Most readers prefer the second one, because it begins simply with a short introductory phrase followed by a one-word subject and a specific verb, then moves toward grammatical complexity.”

5.2 Complex meaning

a. We must first understand cache associativity in order to understand CPU performance. The hit rate, which is the frequency of the data we want residing in the cache, depends on the cache associativity, and CPU performance is better with higher hit rates. The number of cache lines that can be mapped into by a given byte in the RAM gives the associativity. Each cache line is a fixed-sized section of the cache. Although caches with low associativity are easier to design and, on their own, are more efficient than caches with high associativity, caches with high associativity usually have higher hit rates and therefore represent a better choice for good CPU performance.

b. When a CPU needs data from the RAM, it places that data into a small, fast memory on the CPU chip, which we call the cache. We must therefore understand how caches work in order to understand how quickly a CPU can operate. A cache is organized into several cache lines, each of a fixed size. Each byte in the RAM can map into a select set of cache lines, and we call the size of each such set the cache’s associativity. Caches with low associativity are easier to design and, on their own, more efficient, than caches with high associativity. But when we examine how often the data we want resides in the cache—the hit rate—we find that caches with high associativity usually have higher hit rates, thereby allowing the CPU to run faster.

Those passages differ in two ways. First, information that is only implicit in (a) is stated explicitly in (b). More important, note how almost all the technical terms in (a) are toward the beginnings of their sentences and the familiar ones are toward the end.
5.3 Stress

“If you have managed your subjects and topics well, you will, by default put the words you want to emphasize toward the end of your sentences. To test this, read your sentence aloud, and as you reach the last three or four words, tap your finger hard as if emphasizing them in a speech. If you tap on words that do not deserve strong emphasis, look for words that do. Then put those words closer to the end.’’

a. The classes P and NP differ, most theoreticians believe.

b. Most theoreticians believe that the classes P and NP differ.

c. Most theoreticians believe that P doesn’t equal NP.