6

The Funnel: Connecting O and C

Our task, your task . . . is to try to connect the dots.

—Donald Rumsfeld

The opening of a paper identifies a large problem, while the challenge defines a specific question. The main body of the Introduction must connect these elements. It forms the funnel in the hourglass; it narrows the focus and leads readers from the general to the specific, drawing them along the story and framing in the knowledge gap. This is where you build the argument that to make progress on the large problem, you must answer the specific questions.

When you frame the knowledge gap, you provide the background information necessary to understand the story. In an OCAR structure, the background material flows seamlessly from the opening—it is an extension of introducing the problem and the main characters, which is why I don't call it a separate section (hence OCAR, instead of OBCAR). This is in contrast to an ABDCE structure, where after the initial action, you must back up and fill in the background before moving into the development, creating a distinct story element.

Framing the knowledge gap taps into core elements of the SUCCES formula for a sticky story, particularly the U and E elements, unexpectedness and emotion. By defining a knowledge gap, unmasking a hole in the wall of knowledge, you

create unexpectedness: I didn't realize that we didn't know that! By closing with a question, you create curiosity: what is the answer? Then you can tell us how you solve the problem and satisfy our curiosity.

If you do this well, you can bridge from very large problems to very narrow questions. For example, you can argue that to understand the global climate system we need to study bacteria in the frozen soils of the arctic tundra during the winter,[1] or that to cure coronary artery disease in adults it is valuable to map the distribution of ISL1+ cells in early fetal hearts.[2] If you do this badly, expect a rejection letter.

6.1. EXAMPLE OF THE FUNNEL AT WORK

Here is the Introduction from an important paper in atmospheric chemistry.[3] This is an extreme example of narrowing the funnel. It opened with a problem at the global scale (global warming), but the research defined the rate constant of a single chemical reaction. The paper had to convince readers that this extraordinarily constrained piece of laboratory research made a contribution to understanding the global climate system, which it did. That required a careful exercise to connect from the global to the molecular.

In example 6.1, I identify important points with numbers in curly brackets (e.g., {1}), and I eliminated references to make it easier to read the text.

> Example 6.1
> {1} Of all the trace tropospheric species (that is, excluding H_2O and CO_2) methane contributes most to the infrared heating of the atmosphere. {2} Methane is also the most abundant hydrocarbon in the troposphere where it modulates the concentration of the OH free radical and serves as a source of CO. {3} Transport of methane to the stratosphere provides a termination step, via the $Cl + CH_4$ reaction, for the chlorine-catalyzed destruction of ozone. The oxidation of methane in the stratosphere is an important source of water vapour in this region. During the past decade the abundance of methane in the troposphere has been increasing at a rate between 16 and 13 parts per 10^9 volume (p.p.b.v.) per year. {4} The total input and the identities and strengths of the different atmospheric methane sources are not clearly defined. {5} To understand the atmospheric effects of methane, and possibly to regulate it, we need these parameters. {6} At present, the total flux of methane into the atmosphere is estimated from the measured steady-state

1. J. P. Schimel, "The Bugs of Winter: Microbial Control of Soil Biogeochemistry during the Arctic Cold Season," National Science Foundation (2004).

2. Bu et al., "Human ISL1 Hear Progenitors Generate Diverse Multipotent Cardiovascular Cell Lineages," *Nature* 460 (2009): 113–17.

3. G. L. Vaghjiani and A. R. Ravishankara, "New Measurement of the Rate Coefficient for the Reaction of OH with Methane," *Nature* 350 (1991): 406–9.

abundance and the known removal rate of methane. It has been generally accepted that the only process by which methane is chemically degraded in the troposphere is the reaction with OH. {7} Therefore, the rate coefficient, k_1, for the reaction

$$OH + CH_4 \rightarrow CH_3 + H_2O \tag{1}$$

is important in estimating the total flux of methane. The other loss processes, which are expected to be minor pathways, are surface deposition and reaction with Cl atoms in the lower stratosphere and upper troposphere.

{8} A close examination of the available data shows that only in three investigations was k_1 measured below 298 *K*, the temperature region most important to the atmosphere. Only Davis *et al.* measured k_1 down to 240 K. Reaction 1 is slow. Therefore, at low temperatures, the presence of reactive impurities and occurrence of secondary reactions in laboratory systems can result in an overestimate of k_1. {9} We studied reaction 1 using an experimental method in which secondary chemistry could be minimized and the systematic errors reduced.

{1} This is the opening, which frames a story about atmospheric methane (CH_4) and the greenhouse effect. This reaches for a wide audience—it includes anyone interested in global warming, which definitely includes *Nature* editors and readers.

{2} The second sentence introduces the other critical character in this story—OH (hydroxyl radical). By pointing out the CH_4 "modulates" OH, without discussing OH, the authors take for granted that you know why OH is important (a necessary weakness in a paper this short).

{3} This section adds information about why CH_4 is important in the global system. However, because the main story line goes from CH_4 to OH, this material may seem out of place—it goes back to the opening about why CH_4 is important. But the authors presumably felt it important to introduce OH radical as a character early on. Thus, this structure creates an ABDCE story line. The first two sentences formed the A part, and now this backs up to fill in the background (B).

{4} This is a critical statement in laying the base of the knowledge gap: "total input and . . . methane sources are not clearly defined." This paper is going to more clearly define them.

{5} This helps establish the importance of the research—we need to fill the knowledge gap to better manage sources and sinks and mitigate the role of CH_4 in causing global warming.

{6} Here is another critical point in establishing the knowledge gap. The sources of CH_4 are hard to measure, but the major CH_4 sink is reaction with OH radicals, so we can estimate CH_4 fluxes into the atmosphere as being equal to the losses via reaction with OH radicals. This brings the OH radical—a central character—back into play.

{7} At this point, the authors have narrowed all the way down to the molecular scale and the importance of knowing the rate constant for this reaction: to understand the total flux of CH_4 to the atmosphere, we need to know the rate of its reaction with OH, and that means we need to know the rate constant for that reaction. This is essential to understanding the overall role of CH_4 in global warming.

{8} Here, they finish defining the knowledge gap. Having established that we need to know the rate constant k_1, the authors tell us that only three studies have tried to measure k_1 at realistic temperatures, and only one has done so at a temperature that by implication, is in the right range for atmospheric reactions. We need better measurements of k_1 at realistic temperatures to understand atmospheric CH_4 dynamics. These authors quickly scaled down from the global to the micro scale and did so in a way that, at each step, identified what we needed to know. They only tell us what we know to define the limits of that knowledge, rather than for its own sake.

{9} This is the specific statement of the challenge, and unfortunately, I think it's a dud. After clearly framing the knowledge gap and its importance, the authors stated their challenge by saying "We studied Reaction 1 . . ." It's obvious that the question is "What is the value of k_1?" But it would have defined the knowledge gap more concretely to say, "We measured the value of k_1 at temperatures down to 230 *K* using an experimental . . ." This weak challenge highlights an important point: you don't need to be perfect to be successful. This was an important paper.

This was a *Nature* paper and so quite condensed—it had to narrow quickly with broad strokes. In papers for specialist journals, you have more space to develop the Introduction and can do the narrowing more gently and thoroughly. You probably won't have as imposing a task either, narrowing all the way from global to molecular scales. However, the stepwise narrowing process will be the same—make sure that you aren't telling us everything you know about a topic but developing the logical connections between each step to frame the knowledge gap.

6.2. BAD INTRODUCTIONS: FAILING TO DEFINE THE PROBLEM

A good Introduction defines a problem and narrows to an interesting question. A weak or poor Introduction, in contrast, either fails to define the problem or tries to sell a solution before defining the problem, and so fails on curiosity.

6.2.1. Failing to Identify the Problem

Many papers are unclear in defining the problem. They introduce it, tell us that "little is known about this topic," give us some information about it, and close the Introduction by saying "our objectives were to carry out the following tasks." Such works are common in an editor's "New Submissions" folder but are much less frequent in her "Accepted" folder.

The problem with this style of Introduction is that it does a poor job of defining the problem or the value of the solution. It's not very convincing to say "little is known about X" for scientific, logical, and literary reasons.

Scientifically, it is unconvincing because it's probably false. Very few of us have written a paper on a topic that hasn't had tens or hundreds of studies already published on it. Invariably, we know a *lot* about the topic at hand. There are important questions remaining, which is why we did the work, but those are bounded and defined by a large body of knowledge. So when someone says, "little is known about X," we often feel that the author either doesn't know the literature or is overstating the case.

Logically, it's unconvincing because after saying "little is known," the authors describe a lot that *is* known. Even if the short list of facts is everything known on the subject, it comes across as a data dump that contradicts the argument. We don't see the "little."

Finally, linguistically, it's not convincing because it's not concrete. "Little is known" is fuzzy—how little is little? If you tell us the six things that are known, is that still a "little?" Because the language is fuzzy, the argument is unconvincing. To make it convincing, it needs to be concrete—what specifically do we not know?

You must explicitly define the problem, as illustrated in example 6.1. They didn't say that "little is known about CH_4 sources." That would have been inaccurate; we knew a lot about CH_4 sources. Rather, they said "sources are not clearly defined," which is tighter language that implies something closer to "while the broad patterns are known, important details are not," a true description of the state of knowledge at the time and enough to get the paper into *Nature*. A concrete statement that defines a small knowledge gap will do better than a fuzzy one that fails to define one.

6.2.2. Offering a Solution before Defining a Problem

Sometimes authors offer a solution before defining the problem. As you are working on a paper, you live with the topic so closely for so long that it is easy to assume that the question is obvious. It can become hard to see that you haven't posed it clearly. As a result, authors sometimes end up taking the problem for granted and focus on their solution. This creates what I call the "bizzwidget problem" as illustrated by a scenario with a door-to-door salesman: "Hi, ma'am, I'm selling the new Buzco Bizzwidget. The Bizzwidget is the most amazing tool you've ever seen—why, I don't know how you've ever lived without it! So here, let me show you some of the wonderful things it does."

Ma'am is already trying to get rid of the salesman—without finding out that the Bizzwidget really is an amazing tool that she might want to buy. He's trying to sell her a solution, but she doesn't know she has a problem to solve. This strategy works with customers who are inordinately patient, but mostly with people in the "Bizzwidget community" who already know how wonderful the product is.

For the rest of us, we're with the hapless "customer," and the salesman is on the street staring at a closed door. If you are trying to sell us a bizzwidget solution, first convince us we have a problem: "Hi, ma'am—have you experienced problem X? You have? Do you have a solution? You don't? Well I do—let me show it to you; we call it the Buzco Bizzwidget."

Now the Bizzwidget isn't mumbo jumbo. Importantly, this approach engages anyone who has ever experienced problem X, not just the few who have heard of the Bizzwidget. It does this by opening with a concern many people share (defining the audience in the opening), and then *showing* us why we need a Bizzwidget (the body of the Introduction), before introducing the specific product (the challenge). This approach engages our curiosity—do you have a solution? How does it work? It is also concrete—it identifies a real problem and its solution.

If you don't recognize the bizzwidget problem in science writing, consider the following example.

> Example 6.2:
> Addressing complex interactions among chemistry, physics, and biology in climate systems requires an interdisciplinary approach. We propose to address this challenge by using Complex Systems Modeling Theory (CSMT). CSMT has been used in chemical systems to model molecular reaction mechanisms and in cell biology to model physiological pathways. It has been used . . .

This uses the bizzwidget approach; it assumes that we know the problem that CSMT is the solution to and so doesn't define it. It doesn't describe the complex interactions, how other approaches have struggled with them, what the CSMT approach is, or why it is better than other approaches. We may find all that out later in the paper—that CSMT is a solution to a problem we care about—but unless our neighbor already has told us about CSMT, we've probably closed the door on this one. To sell us a solution, first sell us a problem.

6.3. INTRODUCTION VERSUS LITERATURE REVIEW

The need to narrow the focus and lead the reader to your specific questions means that an effective Introduction cannot be merely a literature review that synopsizes what we know about a topic. Instead, because you must convince us of the importance of the problem, you must show us what we *don't* know and why it is important.

The difference between a literature review and an Introduction can be subtle, because they both use the existing literature to discuss the state of knowledge. The distinction between them is that a literature review builds a solid wall—describing knowledge—whereas an Introduction focuses on the hole in that wall—describing ignorance. They tell different stories and move the story in different ways. They also use the existing literature differently; an Introduction focuses on the publications that define the edges, rather than the core of knowledge.

How do tell when you are writing a literature review rather than an Introduction? See whether you are focusing on telling us what we know or what we don't. When you describe something we know, do you use it to identify the boundaries of that knowledge? If so, you're writing an Introduction; if not, you're probably creating a literature review.

One clear flag for when you're doing a literature review is when your citations are at the beginning of sentences. Do you write: "Smith (2003) found X" or do you write: "X occurs (Smith 2003)"? The former tells a story about Smith and what she did; the latter, about nature and how it works. If you write the former, you are probably doing a data dump, collecting the information that seems relevant and writing it down, without synthesizing it and integrating it into a story or framing a knowledge gap. The important information is almost never that Smith found it; rather, it is almost always *what* she found. So why make Smith the subject of the sentence? Whenever you see that you've written a "Smith found . . ." sentence, ask whether the researcher, rather than the research, is what you want to tell us about. If not, rewrite it to focus on the findings. Doing this will help you tighten up the arguments and sharpen the knowledge gap.

There are cases in which you might want to highlight the researcher. The first is when you are discussing an ongoing debate: "Although Smith (2003) reported X, Jones (2005) found Y." This highlights that there is no agreed-on truth but a collection of individual opinions. If there is an accepted dogma that one researcher is challenging, however, you would write something like: "While most reports suggest X (e.g., Smith, 2003, Xu 2004), Jones (2005) found the opposite, arguing . . ."

When there are two camps with multiple papers supporting each side, it is probably best to condense and synthesize it all to "There is still uncertainty about the nature of X, with some reports suggesting it is Y (Smith 2003, Xu 2004) and others suggesting it is Z (Arif 2005, Masukawa 2006)."

The "Smith found X" approach to discussing the literature is common; I do it all the time in my early drafts. But it frequently signals that we haven't fully synthesized the information and figured out why we're presenting it. It is a flag that we're still in the data-dump stage and need at least one more major revision.

Most OCAR papers use a simple O $\rightarrow$ C flow in the Introduction, with a smooth funnel from the opening to the challenge to define the knowledge gap. In contrast, most proposals use a structure more akin to ABDCE. They have an opening section that makes the overall case for the work and briefly sketches in the knowledge gap. Then the background sharpens and fills in that sketch to justify the proposal's specific challenge. But that background is not a literature review—it must still be an introduction. The background is *never* a place for a data dump where you tell us everything about the field. If a piece of information does not have a specific and concrete role in moving the story forward, it does not need to be included.

The vital elements of an Introduction are the opening and the challenge. Those are the "dots" that you must connect by filling in the background and forming the funnel. That material has only one purpose: to show a reader why answering your questions is essential to making progress on the overall problem. By the time readers reach the challenge, they should feel that your questions are the obvious ones, even if they had never thought about them before.

EXERCISES

6.1. Analyze published papers

Go back to the papers you've been analyzing. Look at their Introductions and determine whether they frame the knowledge gap effectively. Does the Introduction have a clean funnel that flows from the opening problem to the specific questions?

6.2. Write a short article

Look at your short article and those of your group. Evaluate the funnel part of the Introduction—does it frame the knowledge gap? If not, revise it so that it does.